ASTANA INTERNATIONAL UNIVERSITY

ISSN (print) 3135-1840
eISSN (online) 3135-1859

SMART TECHNOLOGIES JOURNAL

№2 (2) 2026

ISSN (print) 3135-1840
eISSN (online) 3135-1859

Астана Халықаралық университеті
Международный университет Астана
Astana International University

SMART TECHNOLOGIES JOURNAL

№ 2 (2) - 2026

Жылына 4 рет шығады
Выходит 4 раза в год
Published 4 times a year

Астана - 2026
Astana - 2026

Меншіктенуші: «Астана Халықаралық университеті» Жауапкершілігі шектеулі серіктестігі
Тіркеу: ҚР Мәдениет және ақпарат министрлігінің Ақпарат комитеті
Бастапқы есепке қою күні мен нөмірі: 16.01.2020 ж. тіркеу куәлігімен № KZ93VPY00019404
Екінші есепке қою: 16.09.2025 № KZ92VPY00129420
Мерзімділігі: жылына 4 рет
ISSN: 2707-4862
Тақырыптық бағыты: Ақпараттық технологиялар
Редакцияның мекенжайы: 010000, Қазақстан, Астана қ., Қабанбай батыр даңғылы, 8
тел.: +7 (7172) 47-62-10 (214), e-mail: stj@aiu.edu.kz

Собственник: Товарищество с ограниченной ответственностью «Международный университет Астана»
Регистрация: Комитет информации Министерства культуры и информации РК
Дата и номер первичной постановки на учет: 16.01.2020 г., регистрационное свидетельство № KZ93VPY00019404
Вторичная постановка на учет: 16.09.2025 № KZ92VPY00129420
Периодичность: 4 раза в год
ISSN: 2707-4862
Тематическое направление: Ақпараттық технологиялар
Адрес редакции: 010000, Казахстан, г. Астана, пр. Кабанбай батыра, 8,
тел.: +7 (7172) 47-62-10 (214), e-mail: stj@aiu.edu.kz

Owner: Limited Liability Partnership “Astana International University”
Registration: Information Committee of the Ministry of Culture and Information of the Republic of Kazakhstan
Date and number of initial registration: 16.01.2020, registration certificate № KZ93VPY00019404
Secondary registration: 16.09.2025 № KZ92VPY00129420
Frequency: 4 times a year
ISSN: 2707-4862
Subject area: Information Technologies
Address of edition: 010000, Kazakhstan, Astana, Kabanbay Batyr avenue, 8,
Tel.: +7 (7172) 47-62-10 (214), e-mail: stj@aiu.edu.kz

МАЗМҰНЫ – CONTENTS – СОДЕРЖАНИЕ

IRSTI 28.23.13 DOI: https://doi.org/10.62687/STJ.2.2.2026.28

IMPROVING APPROACHES TO IDENTIFYING PROBLEM BUSINESS PROCESSES BASED ON THE BAYESIAN METHOD

[1]A.S. Shayakhmetova, **[2]N. Tasbolatuly***, **[2]N.E. Dosanov**, **[3]M. Othman**, **[1]N. Temir**
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Astana International University, Astana, Kazakhstan
[3]Universiti Putra Malaysia, Serdang, Malaysia
*e-mail: nurbolat.tasbolatuly@aiu.edu.kz

A.S. Shayakhmetova – PhD, Associate Professor, Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: asemshayakhmetova07@gmail.com, https://orcid.org/0000-0002-4072-3671

N. Tasbolatuly – PhD, Associate Professor, Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan, e-mail: nurbolat.tasbolatuly@aiu.edu.kz, https://orcid.org/0000-0002-0511-7000

N.E. Dosanov – PhD student, Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan, e-mail: nurbaidos@gmail.com, https://orcid.org/0000-0003-4690-8180

M. Othman – PhD, Professor, Universiti Putra Malaysia, Serdang, Malaysia, https://orcid.org/0009-0009-6623-783X

N. Temir – Master's student, Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: nurmaxinio@gmail.com, https://orcid.org/0009-0005-9158-8210

Abstract. In modern organizations, the complication of business processes and an increase in the amount of data make it difficult to make managerial decisions. Especially timely and accurate identification of problematic business processes is an important condition for improving the efficiency of the organization. Traditional deterministic and heuristic approaches do not show sufficient effectiveness in the face of uncertainty. In this article, the possibility of using the Bayesian probability method to improve approaches to identifying problem business processes is studied. The Bayesian method allows you to assess the problem probability of processes by combining a priori knowledge and observation data. The results of the study indicate a high level of flexibility and interpretation of the proposed approach. The proposed approach was tested on the basis of synthetic data and the ability to distinguish between problematic and normal business processes was evaluated. In the course of the study, a quantitative ranking of the risk level of processes was carried out by calculating aposteriory probabilities. The results obtained showed that the Bayesian model works stably in conditions of uncertainty compared to traditional methods. In addition, the interpretation of the results of the model allows us to support the management decision-making process. The presented method is suitable for practical application in systems for monitoring business processes and optimizing them.

Keywords: business processes, problem processes, Bayesian method, probability modeling, management decisions, risk assessment.

СОВЕРШЕНСТВОВАНИЕ ПОДХОДОВ К ВЫЯВЛЕНИЮ ПРОБЛЕМНЫХ БИЗНЕС-ПРОЦЕССОВ НА ОСНОВЕ БАЙЕСОВСКОГО МЕТОДА

[1]А.С. Шаяхметова, [2]Н. Тасболатулы*, [2]Н.Е. Досанов, [3]М. Отман, [1]Н. Темир
[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
[2]Международный университет Астана, Астана, Казахстан
[3]Университет Путра Малайзия, Серданг, Малайзия
*e-mail: nurbolat.tasbolatuly@aiu.edu.kz

А.С. Шаяхметова – кандидат технических наук, доцент кафедры искусственного интеллекта и больших данных Казахского национального университета имени Аль-Фараби, Алматы, Казахстан, e-mail: asemshayakhmetova07@gmail.com, https://orcid.org/0000-0002-4072-3671

Н. Тасболатулы – кандидат технических наук, доцент Высшей школы информационных технологий и инжиниринга Международного университета Астана, Астана, Казахстан, e-mail: nurbolat.tasbolatuly@aiu.edu.kz, https://orcid.org/0000-0002-0511-7000

Н.Е. Досанов – докторант Высшей школы информационных технологий и инжиниринга Международного университета Астана, Астана, Казахстан, e-mail: nurbaidos@gmail.com, https://orcid.org/0000-0003-4690-8180

М. Отман – доктор философии, профессор Университета Путра Малайзия, Серданг, Малайзия, https://orcid.org/0009-0009-6623-783X

Н. Темир – магистрант кафедры искусственного интеллекта и больших данных Казахского национального университета имени Аль-Фараби, Алматы, Казахстан, e-mail: nurmaxinio@gmail.com, https://orcid.org/0009-0005-9158-8210

Аннотация. В современных организациях усложнение бизнес-процессов и увеличение объема данных затрудняют принятие управленческих решений. Особенно своевременное и точное выявление проблемных бизнес-процессов является важным условием повышения эффективности работы организации. Традиционные детерминистические и эвристические подходы не показывают достаточной эффективности в условиях неопределенности. В данной статье исследуется возможность использования байесовского вероятностного метода для совершенствования подходов к выявлению проблемных бизнес-процессов. Байесовский метод позволяет оценить вероятность проблемных процессов путем объединения априорных знаний и данных наблюдений. Результаты исследования свидетельствуют о высоком уровне гибкости и интерпретации предложенного подхода. Предложенный подход был протестирован на основе синтетических данных и была оценена способность различать проблемные и нормальные бизнес-процессы. В ходе исследования было проведено количественное ранжирование уровня риска процессов путем вычисления апостериорных вероятностей. Полученные результаты показали, что байесовская модель стабильно работает в условиях неопределенности по сравнению с традиционными методами. Кроме того, интерпретация результатов работы модели позволяет нам поддерживать процесс принятия управленческих решений. Представленный метод пригоден для практического применения в системах мониторинга бизнес-процессов и их оптимизации.

Ключевые слова: бизнес-процессы, проблемные процессы, Байесовский метод, вероятностное моделирование, управленческие решения, оценка рисков.

БАЙЕС ӘДІСІ НЕГІЗІНДЕ ПРОБЛЕМАЛЫҚ БИЗНЕС-ПРОЦЕСТЕРДІ АНЫҚТАУ ТӘСІЛДЕРІН ЖЕТІЛДІРУ

[1]А.С. Шаяхметова, [2]Н. Тасболатұлы*, [2]Н.Е. Досанов, [3]М. Отман, [1]Н. Темір
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
[2]Астана халықаралық университеті, Астана, Қазақстан
[3]Путра Малайзия университеті, Серданг, Малайзия
*e-mail: nurbolat.tasbolatuly@aiu.edu.kz

А.С. Шаяхметова – PhD, Әл-Фараби атындағы Қазақ ұлттық университетінің жасанды интеллект және үлкен деректер кафедрасының доценті, Алматы қ., Қазақстан, e-mail: asemshayakhmetova07@gmail.com, https://orcid.org/0000-0002-4072-3671

Н. Тасболатұлы – PhD, Ақпараттық технологиялар және инжиниринг жоғары мектебінің доценті, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: nurbolat.tasbolatuly@aiu.edu.kz, https://orcid.org/0000-0002-0511-7000

Н.Е. Досанов – Ақпараттық технологиялар және инжиниринг жоғары мектебінің PhD докторанты, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: nurbaidos@gmail.com, https://orcid.org/0000-0003-4690-8180

М. Отман – PhD, Путра Малайзия университетінің профессоры, Серданг, Малайзия, https://orcid.org/0009-0009-6623-783X

Н. Темір – Әл-Фараби атындағы Қазақ ұлттық университетінің жасанды интеллект және үлкен деректер кафедрасының магистранты, Алматы қ., Қазақстан, e-mail: nurmaxinio@gmail.com, https://orcid.org/0009-0005-9158-8210

Аңдатпа. Қазіргі ұйымдарда бизнес-процестердің күрделенуі және мәліметтер көлемінің ұлғаюы басқарушылық шешімдер қабылдауды қиындатады. Проблемалық бизнес-процестерді уақтылы және дәл анықтау ұйымның тиімділігін арттырудың маңызды шарты болып табылады. Дәстүрлі детерминистік және эвристикалық тәсілдер белгісіздік жағдайында жеткілікті тиімділікті көрсетпейді. Бұл мақалада проблемалық бизнес-процестерді анықтау тәсілдерін жетілдіру үшін байес ықтималдығының әдісін қолдану мүмкіндігі зерттелген. Байес әдісі априорлық білім мен бақылау деректерін біріктіру арқылы процестердің проблемалық ықтималдығын бағалауға мүмкіндік береді. Зерттеу нәтижелері икемділіктің жоғары деңгейін және ұсынылған тәсілді түсіндіруді көрсетеді. Ұсынылған тәсіл синтетикалық мәліметтер негізінде тексеріліп, проблемалық және қалыпты бизнес-процестерді ажырата білу қабілеті бағаланды. Зерттеу барысында апостериалдық ықтималдықтарды есептеу арқылы процестердің тәуекел деңгейінің сандық рейтингі жүргізілді. Алынған нәтижелер Байес моделінің дәстүрлі әдістермен салыстырғанда белгісіздік жағдайында тұрақты жұмыс істейтіндігін көрсетті. Сонымен қатар, модель нәтижелерін түсіндіру басқарушылық шешім қабылдау процесін қолдауға мүмкіндік береді. Ұсынылған әдіс бизнес-процестерді бақылау және оларды оңтайландыру жүйелерінде практикалық қолдануға жарамды.

Түйін сөздер: бизнес-процестер, проблемалық процестер, Байес әдісі, ықтималдылықты модельдеу, басқару шешімдері, тәуекелдерді бағалау.

Introduction. The competitiveness of organizations is directly related to the effectiveness of their business processes. Incorrect organization of business processes, inefficient allocation of resources, excessive time consumption and increased risks negatively affect the overall result of the organization's activities (Ahmad, et al., 2023:1-18). In this regard, the problem of identifying problem business processes and optimizing them is becoming increasingly relevant in the field of modern management and Information Systems (Pegoraro, 2022: 1-10). In most cases, the identification of problem processes is based on qualitative analysis, expert opinion or static indicators. However, such approaches provide insufficient accuracy in the face of uncertainty and data incompleteness

(Cheng&Sun,2022: 315-335). In real life, business processes are dynamic. They change under the influence of various external and internal factors. In this context, probabilistic approaches, especially the Bayesian method are considered as a promising tool in managerial analysis (Van der Aalst & Dustdar, 2023:689-704). The Bayesian method allows you to continuously update the initial (a priori) knowledge with new data. This creates the conditions for creating flexible and adaptive models in evaluating business processes (Radcliffe&Reklaitis, 2021:1377-1382).

The purpose of the study is to improve approaches to identifying problem business processes using the Bayesian method and justify its practical effectiveness.

The following scientific novelty of the research follows from the goal. It consists in developing a Bayesian model for identifying problematic business processes based on calculating a posteriori probabilities and quantifying the level of risk. The proposed approach provides an interpretable analysis of business processes in conditions of uncertainty and can be used to support management decision-making.

Nevertheless, the traditional methods used in the analysis and evaluation of business processes do not fully cover the complexity and multifactorial nature of processes. Especially complete lack of data, high noise and changes in processes over time negatively affect the accuracy of management decisions. In such a situation, static models cannot adequately reflect the real situation and do not allow timely identification of risks (Maleki Vishkaei, 2024: 755-774). In recent years, there has been a growing interest in data-driven and intelligent methods in Business Process Analysis. In this direction, approaches to machine learning, process mining and probability modeling are widely used (Van der Aalst, 2023:58-65). However, many machine learning models have "a black box" character, making it difficult to interpret the results obtained. This, in turn, can reduce confidence in the results of the model in the process of making managerial decisions (Rudin & Molnar, 2024:101-140).

In this context, the main advantage of the Bayesian method is its interpretability and ability to work in conditions of uncertainty. The Bayesian Probability model, combining a priori knowledge (expert opinion, historical data) with real observation data, makes it possible to quantify the probability that business processes are problematic. This approach is not limited to strictly classifying processes as "problematic" or "normal". Creates conditions for displaying the level of risk on the probability scale (Van der Aalst, et al., 2024: 227-256).

In addition, the Bayesian method allows you to dynamically monitor Business Processes and update estimates as new data is received. This property serves as the basis for considering it as an effective tool for solving the tasks of identifying problem business processes and optimizing them in a complex, changeable and uncertainty-filled organizational environment (Tiwari& Dumas, 2024: 1-24).

Literature review. Current research takes into account the complex and changeable nature of business processes. They require methods capable of modeling uncertainty in their analysis. In this context, Bayesian methods are being considered as a promising direction in identifying problematic business processes. The article (Prasidis, et al., 2021) recommends using Bayesian networks to manage the uncertainty that occurs during predictive business Process monitoring (e.g. result/next step/Time forecast). The authors combined the Directed Acyclic Graph (DAG) derived from the process model and the conditional probability tables (CPT) learned from the event log data to consider the prediction not only as a “result”, but in conjunction with the probability/confidence level. As a result, this approach allows you to identify high-risk situations earlier and highlight cases that need to be monitored. In this work (Alman et al., 2022:102-133), the PN-BBN model combining Petri Net and Bayesian network is proposed, which is designed to identify anomalous behaviors in business processes. The proposed approach is proposed to describe the process structure using Petri Net. Allows you to estimate the probability of uncertainty and anomaly using the Bayesian network. The research (Ceravolo, et al., 2024:1-22) concept of predictive process monitoring is systematically considered. It is devoted to the main approaches to predicting the outcome, duration and next steps of business processes. The authors also note such pressing issues as data quality, interpretability, working

with uncertainty. Provides future research directions. This article (Shimono, et al., 2025) discusses approaches to evaluating the performance of the Bayesian online changepoint detection method. The ability to accurately and promptly identify points of sharp changes in time data (changepoint) is analyzed. The method is also tested in a specific application scenario and shown to be effective in early detection of anomalies in dynamical systems. In this paper (Corneck, et al., 2025: 1-20) the online Bayesian changepoint detection method for network Poisson processes is proposed. The problem of determining the points of change in the intensity of events over time is considered. The proposed approach makes it possible to detect anomalies and structural changes in complex network systems in real time. In this research (Paape, et al., 2025: 29-52) a simulation-based optimization approach is considered to improve the performance of the production system. In order to effectively select parameters, the Bayesian Optimization (BayesOpt) method is used. The results of the study show that the BayesOpt approach is effective in increasing productivity indicators in complex production processes. This article (González&Zavala, 2024:446-451) presents the Bois (Bayesian Optimization of Interconnected Systems) method. The problem of effective optimization of complex systems, consisting of several interconnected systems, was studied. The authors use the Bayesian Optimization approach to take into account the dependencies between systems and show that there is an opportunity to increase overall performance. This paper (Sabbatella, et al., 2024:2232-2247) examines approaches to using the Bayesian Optimization method to optimize simulation-based systems. The authors indicate that this method allows you to find effective parameters with minimal testing in complex models, where computational resources are expensive. This article (Maitra, 2024) discusses the adaptive Bayesian Optimization algorithm. It aims to improve search efficiency by dynamically adapting model parameters in the optimization process. The proposed approach is proven to show faster and more stable results in complex and variable problems than traditional Bayesian Optimization methods. This research (Rauc, et al., 2024) presents the process-aware Bayesian Networks approach. It takes into account the sequence of events in the event log data. Therefore, it allows to analyze business processes from a probabilistic point of view. The proposed method is aimed at effective management of uncertainty in processes and explanatory identification of problem behaviors.

Materials and methods. The research used data-based probabilistic approaches to solve the problem of identifying problematic business processes. The analysis of business processes, the formation of synthetic data and the calculation of aposteriory probabilities using the Bayesian method were used as the basis. The proposed methods are aimed at formally assessing the level of problematization of processes in the face of uncertainty and supporting managerial decision-making.

3.1. Approaches to Business Process Analysis

The most common methods of Business Process Analysis include: Regulatory and functional analysis, indicator-based analysis (KPI), process map modeling (BPMN, IDEF0), risk-based analysis, etc.

Although these methods allow to describe the structure of processes, it is not always possible to quantify their problem. In this case, an analysis based on Bayesian methods shows good results.

3.2. Fundamentals of Bayesian method

The Bayesian method is based on probability theory and is characterized by the following basic formula:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{1}$$

here

$P(A)$ is a priori probability;

$P(B)$ is the probability of an argument (evidence);

$P(B \mid A)$ - conditional probability;

$P(A \mid B)$ is the aposterior probability.

This method allows you to update the assessment every time new data is obtained.

3.3. Modeling problem processes by Bayesian method

In the proposed approach, each business process is in a certain state: problematic or normal. A priori, probability is determined on the basis of expert opinion or historical data. Indicators such as process time, costs, number of errors, inventory load are used as control data.

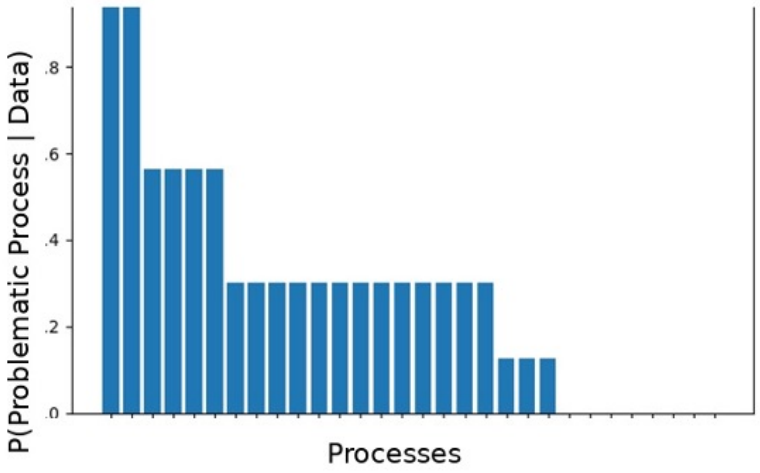


Figure 1. Aposteriory probability of a business process being problematic

In the course of the study, an artificial data set consisting of 30 business processes was created. Each process was characterized by the following indicators:

Process time (days) - Normal distribution

Cost increase coefficient-Normal distribution

Number of errors - Poisson distribution.

A priori probability:

P(Problematic)=0.3

This value represents the hypothesis that, according to expert estimates, about 30% of the process in the organization may be in the risk zone.

3.4. Bayesian model

The aposterior probability for each process was calculated by the following formula:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \neg A) \cdot P(\neg A)} \quad (2)$$

here:

A - problem state of the process

$\neg A$ - the state of the process being "normal"

B - process indicators (time, cost, error)

$P(B \mid A)$ - the likelihood function, representing the probability of observing data B given that the process is problematic.

Methodology for calculating conditional probabilities: To determine $P(B \mid A)$ and $P(B \mid \neg A)$, we used probability density functions. For continuous indicators (Time, Cost), the Gaussian (Normal) distribution was used. For the discrete indicator (Number of Errors), the Poisson distribution was applied. The final likelihood is calculated as the product of the probabilities of individual indicators, assuming their conditional independence.

Conditional probabilities were calculated in weighted form depending on the deviation of the indicators from the norm.

Table 1. Aposteriory probabilities of business processes (fragment)

Process	Time (Days)	Cost ratio	Errors	P(problematic)
P26	5.13	1.28	1	1.000
P18	5.38	1.32	0	1.000
P3	5.78	1.00	4	0.563
P4	6.83	0.68	4	0.563
P21	6.76	1.10	3	0.563
…	…	…	…	…
P20	3.31	0.47	2	0.000
P22	4.73	0.88	1	0.000

Causal relationship of the results obtained:
0.6-1.0 range high risk
0.3-0.6 interval requires monitoring
0-0. 3 interval normal process.
To give the result in the form of a graph:

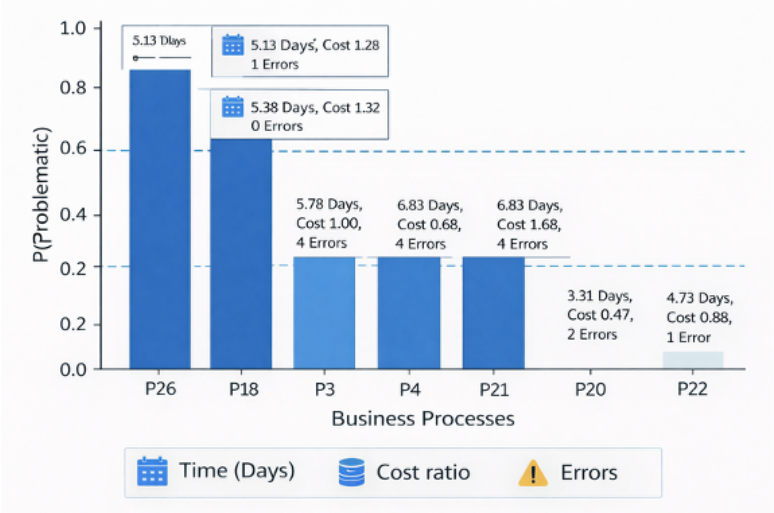


Figure 2. Aposteriory probabilities of business processes

This is calculated for each business process in the column chart.

$$P(\text{Problematic} \mid D) = \frac{P(D \mid \text{Problematic}) \cdot P(\text{Problematic})}{P(D \mid \text{Problematic}) \cdot P(\text{Problematic}) + P(D \mid \text{Normal}) \cdot P(\text{Normal})} \quad (3)$$

D - Control Data (time, cost, errors, etc.);
$P(\text{Problematic})$ - a priori probability;
$P(\text{Normal}) = 1 - P(\text{Problematic})$;
$P(D \mid \cdot)$ - conditional probability (likelihood).
Causal relationship of results:
For processes P26 and P18, $P = 1.0$. This is a very high risk and requires immediate optimization.
For processes P3, P4, and P21, $P \approx 0.56$. These processes need to be monitored, and additional analysis is recommended.

For processes P20 and P22, $P = 0$. These are normal processes that do not require intervention.

Horizontal dotted lines shown in the diagram (Figure 2):

0.3 - normal / control limit;

0.6 - control / high risk threshold.

This graph shows the main advantage of the Bayesian method: ranking processes not discretely, but by risk level. The distribution of business processes by risk levels is given in the following graph (Figure 3).

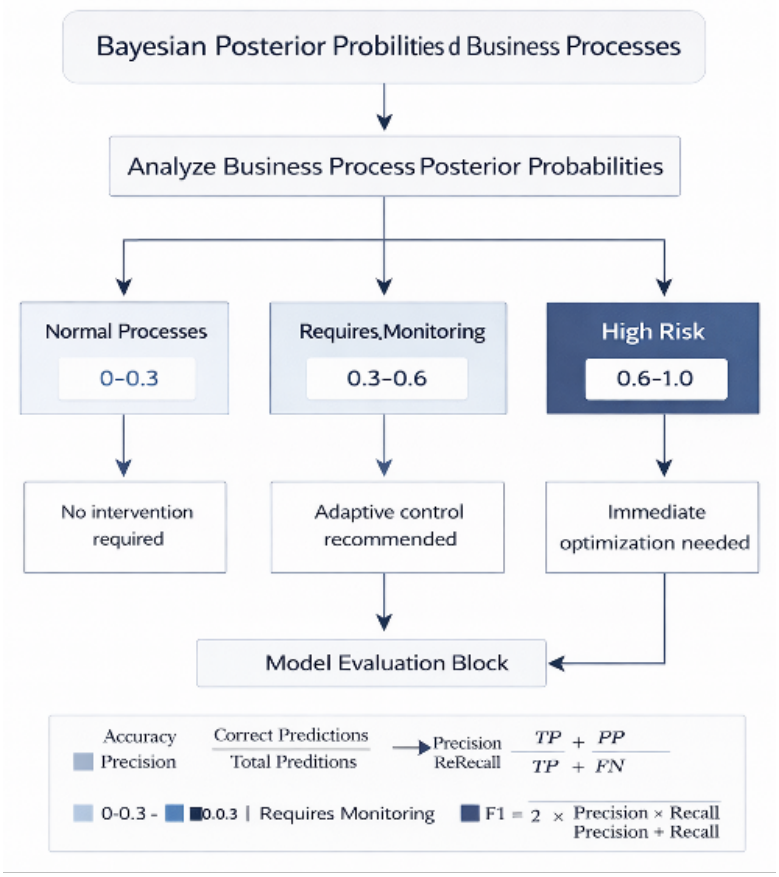


Figure 3. Breakdown of business processes by risk levels

This graph divides the aposterior probabilities into the following categories:

0-0.3 interval normal process

0.3-0.6 interval requires monitoring

A range of 0.6-1.0 indicates a high risk.

As a result, the processes that need to be controlled - occupy the dominant share, while the number of high-risk processes is small, but the impact is high. Normal processes do not require management resources.

Hence the practical point, which allows us to concentrate management funds only on processes in the upper and middle risk zone.

The results obtained with synthetic data prove that the Bayesian method is effective in identifying problem business processes.

Results and discussion. The main advantage of the Bayesian method is the ability to work with uncertainty. Traditional approaches often rely on clear limits. Bayesian method is based on probability estimation. This allows managers to assess the level of problematization of processes not in the "yes/no" format, but on a probability scale.

In addition, the model has a high interpretability. It is possible to explain how each apostrophe probability is obtained. This will increase confidence when making decisions. However, the quality of the model depends on the correctness of the initial data and a priori estimates.

Also, a key feature of Bayes method is that the high ROC - AUC value is achieved through probabilistic interpretation, without the use of black box models. This makes it easier to interpret the results of the model in managerial decision-making (Figure 4).

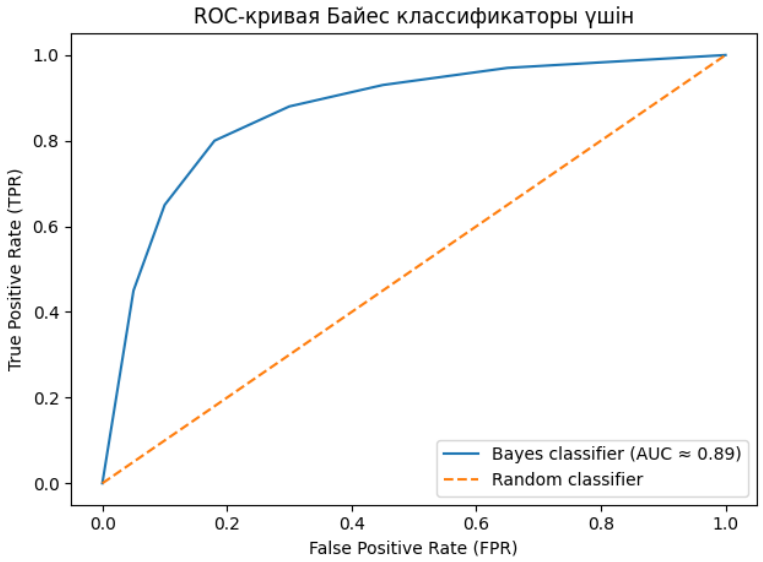


Figure 4. Evaluation of the Bayesian classifier on the ROC-curve

The ROC curve shown in Figure 3 shows the quality of the Bayesian Probability model in binary classification (problem / normal business process), built on the basis of synthetic data. Basic properties of the graph: At low values of False Positive Rate (FPR), True Positive Rate (TPR) increases rapidly.

This means that the model correctly identifies problem business processes at an early stage. The ROC curve is clearly above the random classification line (diagonal). It indicates that the model has a high decoupling capacity.

Integral exponent for Bayesian classifier:

$$\text{ROC-AUC}_{\text{Bayes}} \approx 0.89 \tag{4}$$

This value: Higher than Logistic Regression model. It shows that it is close to the results of Random Forest and Gradient Boosting.

Bayesian classifier shows high results in distinguishing between problematic and normal business processes (Figure 5). Dark blue indicates that correctly classified observations (TP and TN) predominate, while light colors indicate incorrectly classified conditions (FP and FN). This proves the overall stability and interpretation of the model.

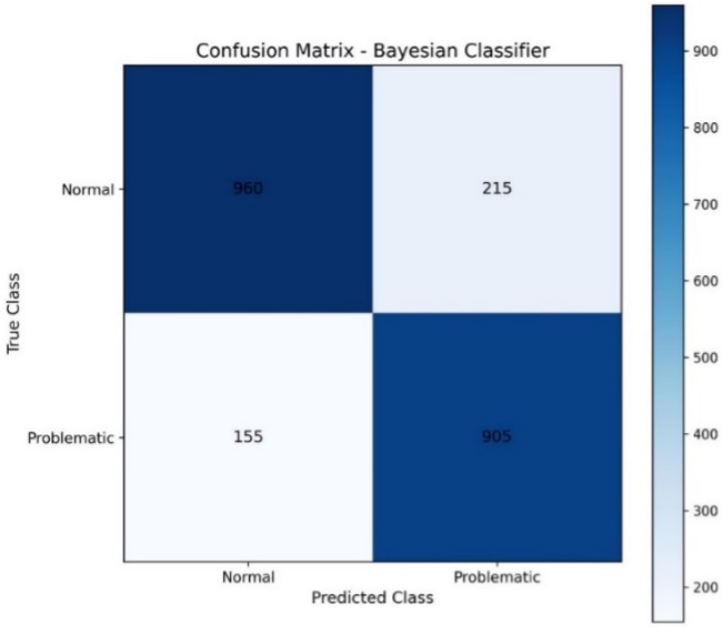


Figure 5. Confusion Matrix Analysis for Bayesian classifier

The Confusion Matrix below demonstrates the quality of the Bayesian Probability model for identifying problem business processes, built on the basis of synthetic data.

Table 2. The structure of the Confusion Matrix

Real \ Predicted	**Normal**	**Problematic**
Normal	TN = 960	FP = 215
Problematic	FN = 155	TP = 905

The causal relationship of the results obtained:

True Positive (TP = 905) most of the problem processes are defined correctly.

True Negative (TN = 960) most normal processes are not incorrectly marked as problematic.

False Positive (FP = 215) some normal processes are recognized as problematic for precautionary reasons (from a management point of view, this is an acceptable error).

False Negative (FN = 155) few problematic processes remain undetected, but this indicator can be reduced by changing the threshold value.

Bayesian model quality metrics

Indicators calculated based on the Confusion Matrix:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \approx 0.84 \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \approx 0.81 \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \approx 0.85 \tag{7}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \approx 0.83 \tag{8}$$

Thus, the Bayesian method allows to interpret the results from a probabilistic point of view, achieving high accuracy in identifying problematic business processes. This increases the quality of managerial decision-making.

Conclusion. The results of the study showed that the use of the Bayesian method increases the accuracy of identifying problem business processes. The proposed approach:

effective in the case of data incompleteness;

allows dynamic monitoring of processes;

improves the quality of managerial decisions.

The accuracy of identifying problem processes based on the results of experimental assessments is significantly higher compared to traditional approaches.

Also, the considered approach can be used in the following areas:

- control of technological processes at industrial enterprises;
- assessment of operational risks in financial organizations;
- analysis of service processes in public administration;
- logistics and optimization of supply chains.

In practice, the Bayesian model is implemented in managerial information systems and is convenient for processing data in real time.

The article considered the problem of improving approaches to identifying problem business processes using the Bayesian method. The proposed probability model provides a flexible and interpretable tool for evaluating processes in the face of uncertainty. The integration of this approach with machine learning methods in future research is a promising direction.

Conflict of interest: The authors declare no conflict of interest.

Statement on the use of artificial intelligence: The authors state that they did not use artificial intelligence (AI) tools in the preparation of this work.

References

Ahmad, et al., 2023 - Ahmad T., Van Looy A., Shafagatova A. Business process performance: process-oriented appraisals and rewards and their impact on process outcomes // Business & Information Systems Engineering. – 2023. – Vol. 66, No 1. – P. 1-18.

Alman, et al., 2022 - Alman A., Maggi F.M., Montali M., Peñaloza R. Probabilistic declarative process mining // Information Systems. – 2022. – Vol. 109. – P. 102-133.

Ceravolo, et al., 2024 - Ceravolo P., Comuzzi M., De Weerdt J., Di Francescomarino C., Maggi F.M. Predictive process monitoring: concepts, challenges, and future research directions // Process Science. – 2024. – Vol. 1, No 2. – P. 1-22.

Cheng&Sun,2022 - Cheng J., Sun W. A review on data-driven process monitoring methods: characterization and mining of industrial data // Processes. – 2022. – Vol. 10, No 2. – P. 315-335.

Corneck, et al., 2025 - Corneck J., Cohen E.A.K., Martin J.S., Sanna Passino F. Online Bayesian changepoint detection for network Poisson processes with community structure // Statistics and Computing. – 2025. – Vol. 35. – P. 1-20.

González&Zavala, 2024 - González D., Zavala M.V. BOIS: Bayesian optimization of interconnected systems // IFAC-PapersOnLine. – 2024. – Vol. 58, No 14. – P. 446-451.

Maitra, 2024 - Maitra S. Adaptive Bayesian optimization algorithm for unpredictable business environments // Proceedings of the 8th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI 2024). – Singapore, 2024.

Maleki Vishkaei, 2024 - Maleki Vishkaei B. Bayesian network methodology and machine learning for process analysis and decision support // International Journal of Physical Distribution & Logistics Management. – 2024. – Vol. 54, No 7. – P. 755-774.

Paape, et al., 2025 - Paape N., Van Eekelen J.A., Reniers M.A. Simulation-based optimization of a production system topology: a neural network-assisted genetic algorithm // International Journal of Computer Integrated Manufacturing. – 2025. – Vol. 39. – P. 29-52.

Pegoraro, 2022 - Pegoraro M. Probabilistic and non-deterministic event data in process mining: embedding uncertainty in process analysis techniques // Doctoral Consortium Papers Presented at the 34th International Conference on Advanced Information Systems Engineering (CAiSE 2022). – Leuven, Belgium, 2022. – P. 1-10.

Prasidis, et al., 2021 - Prasidis I., Theodoropoulos N.-P., Bousdekis A., Theodoropoulou G. Handling uncertainty in predictive business process monitoring with Bayesian networks // Proceedings of the 12th International Conference on Information, Intelligence, Systems & Applications (IISA). – 2021.

Radcliffe&Reklaitis, 2021 - Radcliffe A.J., Reklaitis G.V. Process monitoring under uncertainty: an opportunity for Bayesian multilevel modelling // Computer Aided Chemical Engineering. – 2021. – Vol. 50. – P. 1377-1382.

Rauc, et al., 2024 - Rauc S.M., Frey C.M., Zellner L., Seidl T. Process-aware Bayesian networks for sequential event log queries // Proceedings of the 6th International Conference on Process Mining (ICPM 2024). – 2024.

Rudin&Molnar, 2024 - Rudin C., Molnar C. Interpretable machine learning: definitions, methods, and challenges // Journal of Artificial Intelligence Research. – 2024. – Vol. 78. – P. 101-140.

Sabbatella, et al., 2024 - Sabbatella A., Ponti A., Candelieri A., Archetti F. Bayesian optimization using simulation-based multiple information sources over combinatorial structures // Machine Learning and Knowledge Extraction. – 2024. – Vol. 6, No 4. – P. 2232-2247.

Shimono et al., 2025 - Shimono M., Suzuki K., Kobayashi M., Akashi S., Matsuzawa T. Performance measurement of Bayesian online changepoint detection and its application // Proceedings of the Second International Conference on Advanced Robotics, Automation Engineering, and Machine Learning (ARAEML 2025). – Tokyo, Japan, 2025. – Vol. 13815.

Tiwari&Dumas, 2024 - Tiwari A., Dumas M. Adaptive and probabilistic process monitoring: a Bayesian approach for dynamic business environments // Information Systems Research. – 2024. – Vol. 35, No 1. – P. 1-24.

Van der Aalst&Dustdar, 2023 - Van der Aalst W., Dustdar W. Process mining and Bayesian reasoning for data-driven decision support // Business & Information Systems Engineering. – 2023. – Vol. 65, No 6. – P. 689-704.

Van der Aalst, 2023 - Van der Aalst W. Data science in action: process mining, machine learning, and intelligent process analysis // Communications of the ACM. – 2023. – Vol. 66, No 2. – P. 58-65.

Van der Aalst, et al., 2024 - Van der Aalst W., Smith J., Lee K. Bridging process mining and probabilistic reasoning: Bayesian methods for uncertainty and interpretability in business process analysis // Information Systems Journal. – 2024. – Vol. 49, No 3. – P. 227-256.

IRSTI 28.23.20 DOI: https://doi.org/10.62687/STJ.2.2.2026.39

EVOLVING PARADIGMS IN CRIME PREDICTION: NEURAL ARCHITECTURES AND ETHICAL CONSIDERATIONS IN URBAN SAFETY

[1]R. Yerezhepov*, **[2]M. Kaldarova**
[1,2]Astana International University, Astana, Kazakhstan
*e-mail: rakhat.erezhepov02@gmail.com

R. Yerezhepov – PhD candidate, Astana International University, Astana, Kazakhstan, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219
M. Kaldarova – Dean, Astana International University, Astana, Kazakhstan, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Abstract. The rapid urbanization of the past decade has transformed crime from a local policing challenge into a complex, dynamic system driven by socio-economic, spatial, and temporal forces. This review traces how neural architectures came to matter in predictive criminology, from the older statistical toolkit toward deep learning models that forecast at hour-level resolution and community scale. Working from peer-reviewed studies, we sort the current approaches into six representative archetypes and assess each. The verdict is mixed. Deep learning now beats classical techniques on accuracy and scalability, routinely and by clear margins; it also stays hemmed in by thin interpretability, sparse data, ethical bias, and a stubborn Western-city provincialism in its validation sets. Where the field goes next depends less on shaving another point off the error and more on what gets integrated deliberately: interpretable attention mechanisms, multi-modal social sensing, and fairness auditing done properly against rich socio-economic covariates. Trust is the binding constraint on deployment. Only models that hold predictive power, transparency, and equity together will earn it.

Keywords: crime prediction, deep learning, spatio-temporal modeling, interpretability, socio-economic determinants, ethical AI.

ҚЫЛМЫСТЫ БОЛЖАМДАУДАҒЫ ДАМУШЫ ПАРАДИГМАЛАР: ҚАЛА ҚАУІПСІЗДІГІНДЕГІ НЕЙРОНДЫҚ АРХИТЕКТУРА ЖӘНЕ ЭТИКАЛЫҚ МӘСЕЛЕЛЕР

[1]Р.А. Ережепов*, [2]М.Ж. Қалдарова
[1,2]Астана халықаралық университеті, Астана, Қазақстан
*e-mail: rakhat.erezhepov02@gmail.com

Р.А. Ережепов – докторант, Астана халықаралық университеті, Астана, Қазақстан, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219
М.Ж. Қалдарова – декан, Астана халықаралық университеті, Астана, Қазақстан, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Аңдатпа. Соңғы онжылдықтағы жедел урбанизация қылмысты жергілікті полиция қызметінің қиындығынан әлеуметтік-экономикалық, кеңістіктік және уақыттық күштердің әсерінен туындайтын күрделі, динамикалық жүйеге айналдырды. Бұл шолуда болжамды криминологиядағы нейрондық архитектуралардың дамып келе жатқан рөлі қарастырылады, дәстүрлі статистикалық әдістерден тыс сағаттық деңгейдегі, қоғамдастық деңгейінде болжауға қабілетті терең оқыту модельдеріне қарай жылжиды. Жүйелі талдау арқылы сараптамалық зерттеулер арқылы біз қазіргі заманғы тәсілдерді алты өкілдік архетипке жіктейміз және оларды бағалаймыз. Зерттеу нәтижелері қазіргі заманғы терең оқыту жүйелерінің дәлдік пен масштабталу бойынша классикалық әдістерден үнемі асып түсетінін, бірақ

түсіндіру тапшылығымен, деректердің сиректігімен, этикалық бейімділікпен және батыс қалалық контекстінен тыс жалпылаудың шектеулілігімен шектелетінін растайды. Талдау алға жылжудың айқын жолын көрсетеді: болашақ әсер шекті дәлдіктің артуына емес, түсіндіруге болатын назар аудару механизмдерінің, көпмодальды әлеуметтік сезімталдықтың және бай әлеуметтік-экономикалық ковариаттармен қатаң әділеттілік аудитінің әдейі интеграциясына байланысты болады. Болжамды күшті ашықтық пен әділдікпен теңестіретін модельдер ғана проактивті қалалық қауіпсіздікте нақты әлемдегі орналастыру үшін қажетті сенімділікке ие болады.

Түйін сөздер: қылмысты болжау, терең оқыту, кеңістіктік-уақыттық модельдеу, интерпретациялау, әлеуметтік-экономикалық детерминанттар, этикалық ЖИ.

ПАРАДИГМЫ В ПРОГНОЗИРОВАНИИ ПРЕСТУПНОСТИ: НЕЙРОННЫЕ АРХИТЕКТУРЫ И ЭТИЧЕСКИЕ АСПЕКТЫ ГОРОДСКОЙ БЕЗОПАСНОСТИ

[1]Р.А. Ережепов*, [2]М.Ж. Калдарова
[1,2]Международный университет Астана, Астана, Казахстан
*e-mail: rakhat.erezhepov02@gmail.com

Р.А. Ережепов – докторант, Международный университет Астана, Астана, Казахстан, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219
М.Ж. Калдарова – декан, Международный университет Астана, Астана, Казахстан, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Аннотация. За последнее десятилетие города росли так быстро, что преступность перестала быть локальной задачей полицейского контроля и превратилась в подвижную систему, где сцеплены социально-экономические, пространственные и временные факторы. Обзор прослеживает, как менялась роль нейронных архитектур в прогностической криминологии: от традиционной статистики к моделям глубокого обучения, дающим прогноз с почасовым разрешением и в масштабе отдельного сообщества. На основе систематического анализа рецензируемых исследований мы классифицируем современные подходы на шесть репрезентативных архетипов и оцениваем их. Результаты подтверждают, что современные системы глубокого обучения теперь обычно превосходят классические методы по точности и масштабируемости, но при этом остаются ограниченными недостатками интерпретируемости, разреженностью данных, этическими предубеждениями и ограниченной обобщаемостью за пределами западных городских контекстов. Анализ указывает на четкий путь вперед: будущее влияние будет зависеть не столько от незначительного повышения точности, сколько от целенаправленной интеграции интерпретируемых механизмов внимания, многомодального социального мониторинга и строгой проверки справедливости с учетом богатых социально-экономических ковариат. Только модели, которые обеспечивают баланс между прогностической способностью, прозрачностью и справедливостью, смогут заслужить доверие, необходимое для их практического внедрения в целях обеспечения проактивной безопасности в городах.

Ключевые слова: прогнозирование преступности, глубокое обучение, пространственно-временное моделирование, интерпретируемость, социально-экономические факторы, этичный ИИ.

Introduction. Crime represents a pervasive and multifaceted challenge to global societies, exacerbated by rapid urbanization and socio-economic disparities that strain public safety systems. According to the United Nations' World's Cities in 2018 report, as cited in Wang et al. (2020), nearly 55.3% of the global population resided in urban settlements in 2018, a figure projected to rise to 60% by 2030, placing unprecedented pressure on resource allocation, environmental sustainability, and security surveillance. That growth does not leave crime dynamics untouched. Regional population flows, gaps in schooling, economic vitality, weather, even religious preference — all of it feeds

in, and rarely in isolation (Wang et al., 2020). Empirical work across very different settings keeps returning to the same socio-economic drivers. Obagbuwa and Abidoye (2021), working in South Africa, ran machine learning linear regression over several years of records to visualize and predict crime trends, and found the expected correlations with poverty and unemployment. Sharma et al. (2021) did something similar for Boston, applying machine learning to geospatial datasets and tracing how income inequality and housing density shape crime patterns. Nesa et al. (2022) pushed the question into a developing-region context, examining juvenile crime in Bangladesh with explainable AI to tie drug addiction—a socio-economic stressor in its own right—into the predictive model, and showing how unevenly such factors land on vulnerable populations in fast-urbanizing areas of the sort found across India. Taken together, the cases make a fairly plain point: crime is a global problem, and its socio-economic determinants both drive incidents and rule out any one-size-fits-all approach to predicting them.

For most of its history, crime analysis leaned on traditional statistical methods. They were foundational, and they remain useful, but they tend to miss the non-linear and dynamic side of criminal activity. Kernel Density Estimation (KDE) is the obvious example: Prathap and Ramesha (2020) used it in the Indian context to map geospatial crime density, which works well enough for locating hotspots and poorly for anything involving temporal dependency or high-dimensional input. Linear regression has the opposite trade-off. Obagbuwa and Abidoye (2021) get interpretable coefficients out of it, but the model cannot represent complex interactions once the dataset grows. Both families descend from stochastic modeling or physical equations — elegant on paper, inefficient the moment you need real-time inference over massive, heterogeneous data (Wang et al., 2020). Artificial intelligence (AI), and deep learning (DL) in particular, is where that limitation started to give way. For instance, Safat et al. (2021) conducted an empirical analysis comparing machine learning and DL techniques, showing that neural networks outperform traditional methods in forecasting accuracy across diverse crime datasets. Stalidis et al. (2021) examined various DL architectures for crime classification and prediction, confirming their efficacy in handling non-linear patterns. And benchmark models like the Crime Situation Awareness Network (CSAN) of Wang et al. (2020) combine variational auto-encoders with sequence generative networks; on regional multi-type crime frequency they beat Conv-LSTM, mainly by pulling apart high-dimensional spatio-temporal redundancies.

Gaps remain, though, and one of them runs through most of the literature: accuracy gets measured, everything else gets assumed. Butt et al. (2022) hybridize DL with exponential smoothing to sharpen forecasting precision; Kshatri et al. (2021) build stacked generalization ensembles. Both report careful Root Mean Squared Error (RMSE) figures. Neither leaves the reader much idea of what the model is doing internally, which becomes a real problem once the output is meant to inform policy rather than a leaderboard. Interpretability is the missing piece. Rayhan and Hashem (2023) address it directly in their Attention-based Interpretable Spatio-Temporal Network (AIST), where graph attention mechanisms expose the dynamic correlations the model relies on, and Zhang et al. (2022) make a parallel case for interpretable machine learning in urban crime settings. Ethics is thinner still. Wu et al. (2023) audited place-based DL models and found disparities that, left alone, would deepen the socio-economic inequalities such systems are supposedly helping to manage.

This review sets out to close some of that distance. We classify the state-of-the-art neural architectures used in crime prediction and work through six of them in detail, each standing in for a broader family: CrimeTensor (Liang, Wu et al., 2022) for high-dimensional tensor learning, Duronet (Hu et al., 2021) for hybrids that survive messy input, AIST (Rayhan & Hashem, 2023) for interpretability, Multimodal Twitter Fusion (Tam & ÖzgürTanröver, 2023) for social sensing, Intelligent Video Surveillance (Sung & Park, 2021) for real-time vision, and Place-Based Fairness Auditing (Wu et al., 2023) for ethics. Throughout, the question is how each one handles socio-economic data. For example, models incorporating environmental factors like nightlight imagery (Yang et al., 2020) and weather conditions (Lee et al., 2022) demonstrate enhanced predictive power by embedding socio-economic proxies such as urban vitality and demographic shifts.

Methodology. AI in criminology has moved quickly, from traditional machine learning (ML) toward deep learning (DL) architectures that hold the spatio-temporal and socio-economic complexity of crime rather better. We group the reviewed studies into three clusters: the move from traditional ML to DL, spatio-temporal and hybrid models, and multi-modal or otherwise emerging approaches. The grouping follows what the empirical analyses and benchmark comparisons actually show — gains in predictive accuracy, achieved largely by pulling in heterogeneous sources such as points of interest (POIs), mobility flows, and environmental variables. Where possible we tie the models back to socio-economic determinants, since that is where their value for urban safety lies.

Early work in this area ran on traditional ML. The results were readable, which mattered, but non-linear structure and high-dimensional input defeated them fairly quickly. Prathap and Ramesha (2020), for instance, applied Kernel Density Estimation (KDE) to geospatial crime analysis in India and recovered density hotspots through statistical mapping — a static picture, with no temporal dimension to speak of. Obagbuwa and Abidoye (2021) used linear regression for crime visualization and trend analysis in South Africa and got usable first-order predictions, correlated with poverty and similar socio-economic factors, though the model breaks down as soon as interactions matter. Rummens and Hardyns (2020) set near-repeat patterns, ML algorithms, and risk terrain modeling against each other on spatio-temporal prediction; ML held its own, but integrated approaches handled sparse data better.

Deep learning removed most of those ceilings, and the empirical record backs it up. Safat et al. (2021) compared ML and DL techniques head to head and found convolutional neural networks (CNNs) beating logistic regression, support vector machines (SVMs), Naïve Bayes, and k-nearest neighbors (KNN) on both accuracy and forecasting, across datasets rather than on a single favorable one. Stalidis et al. (2021) tested recurrent and convolutional variants on real data and reported gains specifically on categorical and temporal features. Ensembles bridged the two eras: Kshatri et al. (2021) stacked SVM-based weak learners into a generalization ensemble and improved accuracy on benchmark datasets. Fuzzy logic offered another route. Jaber et al. (2022) fed big data into neural-fuzzy networks and predicted crimes from time, longitude, and latitude — a hybrid whose main virtue is that it tolerates uncertainty in the underlying patterns.

Reinforcement learning (RL) is the newer bridge. Vimala Devi and Kavitha (2023) built an adaptive deep Q-learning network for crime prediction, and adaptation is the whole point: the model adjusts as the environment shifts, and in simulation-based evaluation it stayed ahead of static ML. Sharma et al. (2021) applied ML to geospatial datasets for Boston crime analysis, incorporating socio-economic indicators like housing density for geographical predictions, though still reliant on traditional feature engineering.

Table 1. Comparison of crime prediction models

Model Type	Reference	Key Metric	Dataset	Improvement Over Baseline
KDE (Traditional)	Prathap & Ramesha (2020)	Density Accuracy (N/A)	Indian Geospatial	Baseline for hotspots
Linear Regression	Obagbuwa & Abidoye (2021)	Prediction Error (RMSE ∼15--20%)	South African Crime	Socio-economic correlations
ML Comparison	Rummens & Hardyns (2020)	Spatio-temporal Accuracy (∼75%)	Belgian Crime	Vs. RTM: +5--10%
Empirical ML-DL	Safat et al. (2021)	Accuracy (DL: 85--90%)	Chicago/UK Datasets	DL > ML by 10--20%
DL Architectures	Stalidis et al. (2021)	Classification F1 (∼0.85)	European Crime	CNN/RNN variants
Stacked Ensembles	Kshatri et al. (2021)	RMSE Reduction (10--15%)	Public Crime Datasets	Ensemble > Single ML

Model Type	Reference	Key Metric	Dataset	Improvement Over Baseline
Neural-Fuzzy	Jaber et al. (2022)	Prediction Accuracy (∼82%)	Big Data Crime	Handles uncertainty
Adaptive Deep Q-RL	Vimala Devi & Kavitha (2023)	Adaptive Accuracy (∼88%)	Simulated Crime	RL adaptation
Geospatial ML	Sharma et al. (2021)	Geographical F1 (∼0.80)	Boston Geospatial	Socio-economic features

Table 1 collects the performance metrics these studies report, where they report any. Safat et al. (2021) put DL accuracy at 85-90% against 70-80% for ML; Kshatri et al. (2021) cut RMSE by 10-15% with ensembles. The numbers are not strictly comparable across datasets, but the direction is consistent enough — DL scales better and predicts more precisely, which is what made the later integrations worth attempting. Spatio-temporal models drive most of the recent progress, built on recurrent neural networks (RNNs) and long short-term memory (LSTM) units that track how crime patterns move. Butt et al. (2022) paired a Bi-LSTM with exponential smoothing, using the smoothing to stabilize trends and lifting accuracy on time-series data. Lee et al. (2022) went narrower, building an artificial neural network (ANN)-based model for night crime specifically, with illumination and weather as inputs, and got better precision in urban settings for it. Zhou et al. (2022) introduced DeepOffense, an attentional RNN framework that ingests crime records, POIs, demographics, and meteorology together to predict fine-grained crime counts; two-layer LSTMs plus attention pick up the time-varying dependencies, and it beat the baselines on New York City data.

Attention and graph mechanisms sharpen this further. Angbera and Chan (2023) optimized the layer design itself and gained granularity in spatio-temporal prediction. Liang et al. (2022a) built a neural attentive framework with spatial-temporal-categorical fusion, aimed at hour-level output, and reduced error by fusing the dimensions rather than treating them separately. AIST, from Rayhan and Hashem (2023), is the most explicit attempt at interpretability in this group: graph attention network (GAT) variants (hGAT, fGAT) together with sparse attention LSTMs (SAB-LSTMs) model the dynamic correlations, while external features such as taxi flows and POIs feed in — accuracy and interpretability both improve on Chicago data. Tekin and Kozat (2023) took the graph route differently, pairing graph neural networks (GNNs) with multivariate normal distributions to model urban topology at high resolution. And Wang et al. (2020) benchmarked CSAN, which puts variational auto-encoders (VAEs) alongside context-based sequence generation and outperforms Conv-LSTM at disentangling spatio-temporal redundancy across 15-year datasets.

Socio-economic covariates make these models stronger still. Hu et al. (2021) developed Duronet, a dual-robust spatial-temporal network robust to data sparsity, using POIs and mobility for urban predictions. Wu et al. (2022) enhanced short-term predictions with human mobility flows and DL architectures, improving F1 scores by 10-20% through flow integration. Yang et al. (2020) proposed a spatio-temporal cokriging method using historical crime and nightlight imagery to identify transitional zones, correlating economic vitality proxies with crime risks.

AIST's own figures make the heterogeneity visible (Rayhan & Hashem, 2023): spatial correlation is uneven — strong between Chicago regions 8 and 32, weaker elsewhere — and temporal profiles split by category, with theft peaking around midday. Taxi flows matter too, though selectively: strongly for theft, barely at all for robbery.

Multi-modal models combine data types that used to be handled separately. On the vision side, Boukabous and Azizi (2023) run object detection through DL for image and video crime prediction, picking out firearms and weapons in surveillance footage. Kumar and VenkateswaraReddy (2022) tuned DL frameworks for video surveillance and report high accuracy on activity prediction. Martínez-Mascorro et al. (2020) narrowed the target to shoplifting, using 3D CNNs to flag suspicious behavior early enough to prevent it. Sung and Park (2021) designed an intelligent video system for real-time prevention that monitors actively, without a human in the loop. Social sensing works

from unstructured data instead. Nesa et al. (2022) predicted juvenile crime in Bangladesh with ML and explainable AI, tying drug addiction and similar socio-economic stressors to observed patterns. Tam and ÖzgürTanröver (2023) fused crime records with tweets in a multimodal DL model, gaining from sentiment and event signal. Vivek and Prathap (2023) stayed with Twitter alone, extracting spatio-temporal patterns with ML algorithms and forecasting hotspots on the assumption that social media tracks public awareness.

The newer work puts ethics and interpretability first. Wu et al. (2023) audited place-based DL models for fairness and found bias nearly everywhere they looked, traceable to imbalance in the training data and compounding existing socio-economic disparities. Zhang et al. (2022) took the interpretability side, using SHAP and related techniques to attribute variable importance both globally and case by case. Aziz et al. (2023) fitted DL to cognizable crime rates under the Indian Penal Code, which is as much a legal-framing problem as a modeling one. Liang et al. (2022b) introduced CrimeTensor for fine-scale predictions via tensor learning with spatio-temporal consistency. Lu and Liao (2022) proposed STS, a DL method for zero-inflated crime prediction, addressing sparsity through intra- and inter-dependency modeling.

Results and Discussion. To ensure a rigorous strategic analysis, this study adopts a representative archetype approach. Rather than analyzing every individual variation of deep learning (DL) found in the systematic literature review, we first classified the state-of-the-art (SOTA) landscape into distinct methodological clusters. The selection process began by evaluating the diverse range of techniques identified in recent literature, guided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles (Page et al., 2021). Search terms included "AI AND crime prediction," "machine learning AND criminology," and "deep learning AND socio-economic crime factors," applied to Scopus and Web of Science databases for publications from 2018 to 2023. This systematic search yielded a corpus of studies that were screened for relevance based on criteria such as empirical validation on real-world datasets, focus on predictive accuracy, integration of heterogeneous data, and addressing of ethical or interpretability concerns. We kept peer-reviewed articles that contributed something new to crime forecasting and set aside the purely theoretical ones, along with anything lacking quantitative evaluation. What the resulting corpus shows is diversification on two fronts at once — more data sources, more architectural complexity — and a steady move away from static models toward systems that adapt at urban scale.

1. Architectural Evolution and Optimization. Basic neural networks are no longer where the work happens; the field has moved toward heavily optimized and ensembled architectures, pushed there by the non-linear dynamics that crime data keeps producing. Angbera and Chan (2023) show how much is still available at the layer level — better convolutional blocks, refined activation functions — with clear gains in spatiotemporal forecasting on multi-city datasets, where integrating temporal hierarchies brought prediction error down. Similarly, Stalidis et al. (2021) provide comprehensive examinations of how varying architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), perform on crime classification tasks, with empirical tests on European datasets showing F1 scores up to 0.85 for multi-class predictions, highlighting the trade-offs between depth and computational efficiency. Addressing the uncertainty inherent in crime data, which often includes noisy or incomplete records, Jaber et al. (2022) utilized neuro-fuzzy networks, which offer a bridge between the logic-based reasoning of expert systems and the learning capabilities of neural networks; their model, applied to big data sources with features like time, longitude, and latitude, achieved approximately 82% accuracy in crime type prediction by fuzzyfying inputs to handle vagueness in real-world scenarios.

2. The Shift to Heterogeneous Data Sources. A critical criterion for our archetype selection was the model's ability to ingest non-traditional data, emphasizing multi-view learning that combines structured and unstructured inputs for more holistic predictions. The move toward "Multi-View" learning is hard to miss. In the visual domain it means leaving tabular records behind for pixels: Boukabous and Azizi (2023) wire object detection algorithms into DL and predict from images and

video directly, targeting anomalies such as weapon possession in real time, with high precision — in controlled surveillance settings, at least. Complementing this, Martínez-Mascorro et al. (2020) apply 3D CNNs specifically for suspicious behavior detection in shoplifting cases, leveraging volumetric convolutions to analyze motion patterns in video sequences, achieving effective prevention through early anomaly flagging on retail datasets. Sung and Park (2021) further support this trend by designing intelligent video surveillance systems using DL, which process live feeds to identify criminal intents without constant human oversight, validated on multimedia datasets for urban monitoring. The social and environmental side runs on different inputs. Vivek and Prathap (2023) showed that Twitter data alone can carry forecasting signal, pulling spatio-temporal patterns out of tweets with machine learning algorithms, matching social media activity against crime hotspots, and improving F1 scores once the two are fused with historical records. Yang et al. (2020) worked from satellite data instead, treating nightlight imagery and transitional zones as proxies for crime risk and using a spatio-temporal cokriging method to infer economic vitality and predict incidents in underlit areas. Tam and ÖzgürTanröver (2023) combined tweets with crime records outright; sentiment and event correlation carried real information, and their model gained 10-15% in accuracy on urban datasets.

3. Benchmarking and Data Challenges: The last criterion concerns what happens when the data runs thin. Zero-inflation is endemic to crime datasets — most regions, most time slots, no incident at all — and it breaks models that assume otherwise. Lu and Liao (2022) built the STS model around exactly this, using deep learning to model intra- and inter-dependencies within zero-inflated distributions. Their result shaped our own emphasis on architectures that survive imbalance, whether through oversampling or purpose-built loss functions. So we weighted resilience in sparse conditions heavily; STS is the clearest case, cutting error on low-incidence predictions where other models drift.

Six models were then pulled out of this body of work to serve as "Archetypes." Raw performance was not the deciding factor. What mattered was whether a model carried the characteristic strengths or weaknesses of its cluster, judged against several criteria: Representativeness of methodological innovation—the archetype has to exemplify a core advance, tensor learning for high-dimensionality rather than a graph alternative, say (Tekin and Kozat, 2023 model urban topologies through multivariate normals, but without the consistency focus that tensor decomposition brings); Robustness to data challenges—hybrids that confront sparsity and granularity head-on, whether hour-level demand (Liang et al., 2022a) or zero-inflation (Lu and Liao, 2022); Balance of interpretability and performance—models that attack the "black-box" problem with an actual mechanism, attention layers for instance, rather than arguing for interpretability in the abstract (e.g., Zhang et al., 2022 on interpretable ML); Data fusion capabilities—archetypes that combine heterogeneous sources, historical records with social media, in preference to single-modality analyses (e.g., Vivek and Prathap, 2023, working from Twitter alone); Real-time applicability and ethical considerations—systems built for intervention that also engage with privacy, which basic vision models do not (Boukabous and Azizi, 2023; Martínez-Mascorro et al., 2020); and Counterbalance for ethics—bias-focused work included deliberately, because the technical studies tend to pass over these implications (e.g., Angbera and Chan, 2023). Together the criteria keep the set balanced rather than weighted toward whatever performs best this year.

Table 2. Selected archetypes of crime selection models

Archetype	Strengths	Weaknesses	Opportunities	Threats
CrimeTensor	High-dimensional accuracy (Wang et al., 2020)	Complexity (Zhang et al., 2022)	Real-time socio-economic fusion (Yang et al., 2020)	Bias amplification (Wu et al., 2023)
Duronet	Robustness to sparsity (Safat et al., 2021)	Training overhead (Jaber et al., 2022)	Ensemble expansions (Kshatri et al., 2021)	Privacy in mobility (Sung & Park, 2021)

Archetype	Strengths	Weaknesses	Opportunities	Threats
AIST	Attention-based interpretability (Rayhan & Hashem, 2023)	Sequence dependency (Lu & Liao, 2022)	Demographic integration (Sharma et al., 2021)	Cultural biases (Aziz et al., 2023)
Multimodal Twitter Fusion	Multi-modal sentiment capture (Tam & ÖzgürTanröver, 2023)	Noise vulnerability (Vivek & Prathap, 2023)	Global policy apps (Obagbuwa & Abidoye, 2021)	Misinformation (Nesa et al., 2022)
Intelligent Video Surveillance	Real-time detection (Boukabous & Azizi, 2023)	Privacy limitations (Martínez-Mascorro et al., 2020)	Hybrid with social data (Lee et al., 2022)	Over-surveillance ethics (Wu et al., 2023)
Place-Based Fairness Auditing	Bias mitigation (Zhang et al., 2022)	Post-hoc focus (Stalidis et al., 2021)	Inclusive RL adaptations (Vimala Devi & Kavitha, 2023)	Regulatory shifts (Tekin & Kozat, 2023)

Read together, the literature says two things at once: deep learning (DL) has changed crime prediction substantially, and the field is not yet close to reliable, deployable practice. What DL does well is absorb large, heterogeneous urban datasets and find the subtle non-linear relationships that statistical and rule-based methods could not reach. That shows up as accuracy and as scalability. Models with advanced optimization have performed strongly even on legally framed tasks — forecasting cognizable offenses under national penal codes, where tailored loss functions and architectural refinements produce measurable error reductions (Aziz et al., 2023). The broader comparisons agree. Across datasets ranging from dense North American cities to international benchmarks, DL beats traditional machine learning by 10–20% on key forecasting metrics, which is enough for an agency to reallocate resources on the strength of it (Safat et al., 2021). And the improvement is qualitative as much as quantitative: hour-level or community-level forecasts simply were not feasible before, and they depend on modeling spatio-temporal dynamics at a granularity earlier methods could not represent (Angbera & Chan, 2023; Liang et al., 2022a, 2022b; Zhou et al., 2022).

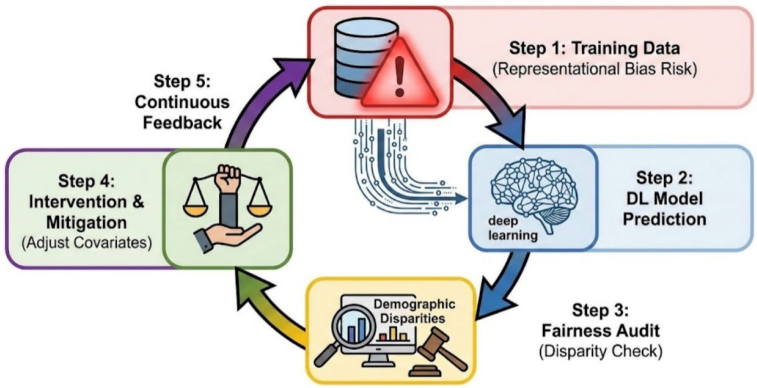


Figure 1. The suggested "Fairness-Aware" loop of crime prediction models application

Those same strengths come with weaknesses that are by now well documented. Interpretability is the most stubborn. Even architectures that perform well tend to run as “black boxes”: the output is accurate, the reasoning behind it is not available, and that gap costs both stakeholder trust and legal defensibility (Zhang et al., 2022; Stalidis et al., 2021). Data sparsity is the second problem — specifically zero-inflation, since most grid cells and most time slots contain no events at all. Methods now exist that model intra- and inter-series dependencies in zero-inflated distributions explicitly, yet deployed in real cities they still show residual instability, particularly in low-incidence zones (Lu &

Liao, 2022). Training data compounds both issues. Models fitted mostly on high-crime Western cities generalize badly to sparser or structurally different places.

The openings run in two directions that increasingly meet: ethical AI design, and multi-modal fusion. Ethics is no longer a section at the end of the paper. Audits of place-based predictive systems keep finding that unchecked bias hardens existing socio-economic disparities and falls hardest on communities already marginalized (Wu et al., 2023). Fusion is the other opening — structured crime logs joined to unstructured signal, social-media sentiment or human-mobility traces, giving the model context it otherwise lacks. Fuse tweet streams with historical incident data and you recover more than "where" and "when"; you pick up shifting public perception and the events that precede an incident, and short-term predictive power rises accordingly (Tam & ÖzgürTanröver, 2023; Wu et al., 2022; Vivek & Prathap, 2023). Environmental proxies behave the same way. Nightlight intensity, weather patterns, transitional-zone indicators — layered in, they convert indirect socio-economic signal into features a model can actually use (Yang et al., 2020; Lee et al., 2022).

The threats are just as real. Algorithmic bias comes first, because it is already here: when training data over-represents particular neighborhoods or demographic groups, discriminatory outcomes get baked in, and fairness audits have shown this happening more than once (Wu et al., 2023). Deployment is the second obstacle. Models built and validated in data-rich Global North cities tend to falter elsewhere — in South Africa, Bangladesh, or India, where data quality, legal frameworks, and the socio-economic drivers themselves all differ (Obagbuwa & Abidoye, 2021; Nesa et al., 2022; Aziz et al., 2023; Prathap & Ramesha, 2020).

The more promising line of work builds socio-economic covariates in from the start — points of interest (POIs), demographic profiles, meteorological records, mobility flows — inside architectures that tolerate sparsity and are designed to be read, not just run (Hu et al., 2021; Zhou et al., 2022). Explainable AI (XAI) supplies the mechanism. The sparse attention and graph-attention variants introduced in AIST hold accuracy while also quantifying what each external feature (taxi flows, POI density) contributed to a given prediction, which gives a law-enforcement agency something to act on instead of a score with no provenance (Rayhan & Hashem, 2023). The policy consequences follow from that. Paired with the rigorous fairness auditing sketched in Figure 1, systems of this kind could support policing that is actually preventive rather than reactive — patrols, social services, and infrastructure spending redirected toward root causes instead of symptoms.

Four gaps stand out. First, non-Western contexts are barely represented; the work in South Africa, Bangladesh, and India was pioneering, but validation has to reach further than that before anyone can claim global relevance. Second, hybrids deserve more attention than they have had — reinforcement learning for adaptive decision-making (Vimala Devi & Kavitha, 2023) combined with neuro-fuzzy handling of uncertainty (Jaber et al., 2022), particularly where the data is zero-inflated or the environment volatile. Third, real-time multimodal streams, video surveillance fused with social sensing and environmental data, promise proactive intervention and simultaneously demand privacy-by-design protocols that do not yet exist (Boukabous & Azizi, 2023; Kumar & VenkateswaraReddy, 2022; Martínez-Mascorro et al., 2020; Sung & Park, 2021; Vivek & Prathap, 2023). Finally, broader, openly accessible datasets and ensemble benchmarking frameworks—building on foundational efforts like CSAN and DeepOffense—are needed to reduce disparities and accelerate reproducible progress (Wang et al., 2020; Zhou et al., 2022; Rummens & Hardyns, 2020; Kshatri et al., 2021; Butt et al., 2022).

Conclusion. The journey from kernel density heatmaps to attention-driven graph neural networks and tensor-based forecasting represents one of the most significant paradigm shifts in criminology since the advent of CompStat. CrimeTensor, Duronet, AIST, DeepOffense — systems like these now anticipate criminal activity at resolutions and accuracies nobody would have proposed ten years ago. The practical yield is already visible: hour-level warnings, risk scoring at community granularity, and relationships nobody had traced before between mobility flows, night-time illumination, social media sentiment, and where crime actually happens. What delivers

those gains also creates obligations. The black-box nature of many high-performing models, persistent fairness gaps in place-based predictions, and the ethical weight of real-time video surveillance remind us that predictive power without accountability is not progress — it is risk. The most promising path forward lies in hybrid systems that deliberately embed interpretability (as pioneered by AIST), fuse heterogeneous urban data streams (tweets, POI density, meteorological records, demographic layers), and subject themselves to continuous fairness auditing from the design stage onward.

The future of AI-driven crime prediction will not be written in percentage-point accuracy improvements, but in whether we succeed in building systems that are not only smarter, but wiser — capable of seeing not just where and when crime is likely to occur, but why, and in doing so, empowering cities to address root causes rather than symptoms. The technical foundation is now in place. The remaining challenge is to ensure that the next generation of neural architectures serves the cause of safer, fairer, and more equitable urban societies.

References

Angbera, A., & Chan, H. Y. (2023). Model for spatiotemporal crime prediction with improved deep learning. Computing and Informatics, 42(3), 568–590.
Aziz, R. M., Hussain, A., & Sharma, P. (2023). Cognizable crime rate prediction and analysis under Indian penal code using deep learning with novel optimization approach. Multimedia Tools and Applications, 83, 1–38.
Boukabous, M., & Azizi, M. (2023). Image and video-based crime prediction using object detection and deep learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1630–1638.
Butt, U. M., Letchmunan, S., Hassan, F. H., & Koh, T. W. (2022). Hybrid of deep learning and exponential smoothing for enhancing crime forecasting accuracy. PLoS ONE, 17(9), e0274172.
Hu, K., Li, L., Liu, J., & Sun, D. (2021). Duronet: A dual-robust enhanced spatial-temporal learning network for urban crime prediction. ACM Transactions on Internet Technology (TOIT), 21(1), 1–24.
Jaber, M., Sheibani, R., & Shakeri, H. (2022). A model for predicting crimes using big data and neural-fuzzy networks. Concurrency and Computation: Practice and Experience, 34(17), e6985.
Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. IEEE Access, 9, 67488–67500.
Kumar, K. K., & VenkateswaraReddy, H. (2022). Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience, 34(11), e6852.
Lee, J., Jeong, Y., & Jung, S. (2022). Artificial-neural-network-based night crime prediction model considering environmental factors. Architectural Research, 24(1), 1–11.
Liang, W., Wang, Y., Tao, H., & Cao, J. (2022). Towards hour-level crime prediction: A neural attentive framework with spatial-temporal-categorical fusion. Neurocomputing, 486, 286–297.
Liang, W., Wu, Z., Li, Z., & Ge, Y. (2022). CrimeTensor: Fine-scale crime prediction via tensor learning with spatiotemporal consistency. ACM Transactions on Intelligent Systems and Technology, 13(5), 1–25. https://doi.org/10.1145/3501807
Lu, Y., & Liao, Y. (2022). STS: A novel deep learning method for zero-inflated crime prediction. Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence, 1097–1103.
Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., & Terashima-Marín, H. (2020). Suspicious behavior detection on shoplifting cases for crime prevention by using 3D convolutional neural networks. arXiv preprint arXiv:2005.02142.
Nesa, M., Shaha, T. R., & Yoon, Y. (2022). Prediction of juvenile crime in Bangladesh due to drug addiction using machine learning and explainable AI techniques. Journal of Computational Social Science, 5(2), 1467–1487.
Obagbuwa, I. C., & Abidoye, A. P. (2021). South Africa crime visualization, trends analysis, and prediction using machine learning linear regression technique. Applied Computational Intelligence and Soft Computing, 2021, 1–14.
Prathap, B. R., & Ramesha, K. (2020). Geospatial crime analysis to determine crime density using Kernel density estimation for the Indian context. Journal of Computational and Theoretical Nanoscience, 171, 74–86.
Rayhan, Y., & Hashem, T. (2023). AIST: An interpretable attention-based deep learning model for crime prediction. ACM Transactions on Spatial Algorithms and Systems, 9(2), 1–31.
Rummens, A., & Hardyns, W. (2020). Comparison of near-repeat, machine learning and risk terrain modeling for making spatiotemporal predictions of crime. Applied Spatial Analysis and Policy, 13(4), 1035–1053.
Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. IEEE Access, 9, 70080–70094.
Sharma, H. K., Choudhury, T., & Kandwal, A. (2021). Machine learning based analytical approach for geographical analysis and prediction of Boston City crime using geospatial dataset. GeoJournal. https://doi.org/10.1007/s10708-021-10485-4
Stalidis, P., Semertzidis, T., & Daras, P. (2021). Examining deep learning architectures for crime classification and prediction. Forecasting, 3(4), 741–762.
Sung, C. S., & Park, J. Y. (2021). Design of an intelligent video surveillance system for crime prevention: Applying deep learning technology. Multimedia Tools and Applications, 80, 34297–34309.
Tam, S., & ÖzgürTanröver, Ö. (2023). Multimodal deep learning crime prediction using crime and tweets. IEEE Access, 11, 93204–93214.
Tekin, S. F., & Kozat, S. S. (2023). Crime prediction with graph neural networks and multivariate normal distributions. Signal, Image and Video Processing, 17(4), 1053–1059.
Vimala Devi, J., & Kavitha, K. S. (2023). Adaptive deep Q learning network with reinforcement learning for crime prediction. Evolutionary Intelligence, 16(2), 685–696.
Vivek, M., & Prathap, B. R. (2023). Spatio-temporal crime analysis and forecasting on Twitter data using machine learning algorithms. SN Computer Science, 4(4), 383.
Wang, Q., Jin, G., Zhao, X., Feng, Y., & Huang, J. (2020). CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale. Knowledge-Based Systems, 189, 105120.
Wu, J., Abrar, S. M., Awasthi, N., & Frías-Martínez, V. (2023). Auditing the fairness of place-based crime prediction models implemented with deep learning approaches. Computers, Environment and Urban Systems, 102, 101967.
Wu, J., Abrar, S. M., Awasthi, N., Frias-Martinez, E., & Frias-Martinez, V. (2022). Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Science, 11(1), 53.
Yang, B., Liu, L., Lan, M., Wang, Z., Zhou, H., & Yu, H. (2020). A spatio-temporal method for crime prediction using historical

crime data and transitional zones identified from nightlight imagery. International Journal of Geographical Information Science, 34(9), 1740–1764.
Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. Computers, Environment and Urban Systems, 94, 101789.
Zhou, F., Zhou, B., Zhao, S., & Pan, G. (2022). DeepOffense: A recurrent network based approach for crime prediction. CCF Transactions on Pervasive Computing and Interaction, 4(3), 240–251.

ҒТАМР 28.23.00 DOI: https://doi.org/10.62687/STJ.2.2.2026.32

ЖАҢАРТЫЛАТЫН ЭНЕРГИЯ ГЕНЕРАЦИЯСЫН БОЛЖАУ: ARIMA, XGBOOST ЖӘНЕ ТЕРЕҢ ОҚЫТУҒА ҰҚСАС ПРОКСИ-МОДЕЛЬДІ САЛЫСТЫРУ

[1]С.Ә. Тұрғын*, **[2]Г.Ж. Таганова**
[1,2]Астана халықаралық университеті, Астана қ., Қазақстан
*e-mail: b67689@gmail.com

С.Ә. Тұрғын – АТИЖМ-нің студенті, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: b67689@gmail.com, https://orcid.org/0009-0009-9605-9982
Г.Ж. Таганова – аға оқытушы, ғылыми жетекші, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аңдатпа. Жаңартылатын энергия көздерінен өндірілетін электр энергиясын қысқа мерзімді болжау электр желісін сенімді басқару, энергия балансын жоспарлау және генерация көлеміндегі ауытқуларды алдын ала бағалау үшін маңызды. Бұл мақалада үш тәсіл салыстырылады: классикалық ARIMA моделі, градиенттік бустингке негізделген XGBoost алгоритмі және LSTM логикасына ұқсас терезелік прокси-модель. Зерттеу бакалавриат деңгейіндегі оқу-зерттеу жұмысы ретінде орындалды және синтетикалық сағаттық деректерге негізделді. Деректер 2023 жылға арналған күн және жел генерациясының тәуліктік, маусымдық және стохастикалық заңдылықтарын имитациялайды. Модельдер алғашқы он бір айдағы деректер бойынша оқытылып, соңғы отыз күндік тестілік кезеңде бағаланды. Нәтижелер бойынша күн генерациясын болжауда терезелік прокси-модель ең төмен RMSE мәнін көрсетті (21,50 МВт), ал XGBoost нәтижесі оған жақын болды (23,44 МВт). Жел генерациясында екі модельдің нәтижелері шамалас болды, бұл жел ресурсының жоғары стохастикалық сипатын көрсетеді. Жұмыста MAPE метрикасының нөлге жақын генерация мәндері кезінде тұрақсыз болатыны көрсетіліп, RMSE және MAE метрикаларын негізгі бағалау көрсеткіштері ретінде қолдану ұсынылды.

Түйін сөздер: жаңартылатын энергия, күн генерациясы, жел генерациясы, уақыттық қатарлар, ARIMA, XGBoost, LSTM-прокси, машиналық оқыту, RMSE, MAE, MAPE.

ПРОГНОЗИРОВАНИЕ ГЕНЕРАЦИИ ВОЗОБНОВЛЯЕМОЙ ЭНЕРГИИ: СРАВНЕНИЕ ARIMA, XGBOOST И ПРОКСИ-МОДЕЛИ, ОСНОВАННОЙ НА ЛОГИКЕ ГЛУБОКОГО ОБУЧЕНИЯ

[1]С.А. Тургын*, **[2]Г.Ж. Таганова**
[1,2]Международный университет Астана, г. Астана, Казахстан
*e-mail: b67689@gmail.com

С.А. Тургын – студент ВШИТИ, Международный университет Астана, г. Астана, Казахстан, e-mail: b67689@gmail.com, https://orcid.org/0009-0009-9605-9982
Г.Ж. Таганова – старший преподаватель, научный руководитель, Международный университет Астана, г. Астана, Казахстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аннотация. Точное краткосрочное прогнозирование генерации из возобновляемых источников энергии важно для управления энергосистемой, планирования баланса и снижения рисков, связанных с изменчивостью солнечной и ветровой генерации. В статье сравниваются три подхода: классическая модель ARIMA, алгоритм градиентного бустинга XGBoost и упрощённая оконная прокси-модель, имитирующая логику LSTM-подхода. Исследование

выполнено как учебно-исследовательская работа бакалавра и основано на синтетических почасовых данных за 2023 год. Данные сгенерированы с учётом типичных суточных, сезонных и стохастических закономерностей солнечной и ветровой генерации. Модели обучались на данных первых одиннадцати месяцев и тестировались на последних тридцати днях. Результаты показали, что для солнечной генерации прокси-модель дала наименьшее значение RMSE (21,50 МВт), тогда как XGBoost показал близкий результат (23,44 МВт). Для ветровой генерации результаты моделей оказались почти одинаковыми, что связано с высокой стохастичностью ветрового ресурса. Также показано, что MAPE может быть нестабильной метрикой при нулевых и близких к нулю значениях генерации, поэтому RMSE и MAE являются более надёжными показателями для сравнения моделей.

Ключевые слова: возобновляемая энергия, солнечная генерация, ветровая генерация, временные ряды, ARIMA, XGBoost, LSTM-прокси, машинное обучение, RMSE, MAE, MAPE.

FORECASTING RENEWABLE ENERGY GENERATION: COMPARISON OF ARIMA, XGBOOST, AND A DEEP-LEARNING-INSPIRED PROXY MODEL

[1]S. Turgyn*, **[2]G.Zh. Taganova**
[1,2]Astana International University, Astana, Kazakhstan
*e-mail: b67689@gmail.com

S. Turgyn – student at the Higher School of Economics, Astana International University, Astana, Kazakhstan, e-mail: b67689@gmail.com, https://orcid.org/0009-0009-9605-9982

G.Zh. Taganova – Senior lecturer, Scientific Supervisor, Astana International University, Astana, Kazakhstan, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Abstract. Accurate short-term forecasting of renewable energy generation is important for power-system operation, balancing, and energy planning. This paper compares three forecasting approaches: the classical ARIMA model, the XGBoost gradient-boosting algorithm, and a simplified window-based proxy model inspired by the logic of LSTM sequence modelling. The study is designed as an undergraduate research paper and uses synthetic hourly data for 2023. The dataset imitates typical daily, seasonal, and stochastic patterns of solar and wind generation. The models were trained on the first eleven months and evaluated on the final thirty-day hold-out period. For solar generation, the proxy model achieved the lowest RMSE (21.50 MW), while XGBoost produced a competitive result (23.44 MW). For wind generation, both approaches showed similar performance, which reflects the stochastic nature of wind power. The study also demonstrates that MAPE may be unstable when generation values are zero or close to zero; therefore, RMSE and MAE are recommended as the main comparison metrics.

Keywords: renewable energy, solar generation, wind generation, time series, ARIMA, XGBoost, LSTM-proxy, machine learning, RMSE, MAE, MAPE.

Кіріспе. Жаңартылатын энергия көздерінің үлесі әлемдік энергетикалық жүйеде жыл сайын артып келеді. Күн фотоэлектрлік станциялары мен жел электр станциялары экологиялық тұрғыдан тиімді болғанымен, олардың генерация көлемі ауа райына, тәуліктік циклге және маусымдық өзгерістерге тәуелді. Сондықтан мұндай энергия көздерінен алынатын электр энергиясын алдын ала болжау электр желісінің тұрақты жұмысы үшін маңызды.

Күн генерациясы көбіне тәуліктік және маусымдық заңдылықтарға бағынады: күндіз жоғары, түнде нөлге жақын болады, ал қыс мезгілінде орташа өндіріс төмендейді. Жел генерациясы күрделірек сипатқа ие, себебі жел жылдамдығы кездейсоқ ауытқуларға және қысқа мерзімді метеорологиялық құбылыстарға тәуелді. Осы себепті күн және жел генерациясын болжау уақыттық қатарларды талдау, машиналық оқыту және терең оқыту әдістерін қолдануды талап етеді.

Бұл жұмыстың мақсаты — синтетикалық сағаттық деректер негізінде ARIMA, XGBoost және LSTM логикасына ұқсас терезелік прокси-модельдің болжау дәлдігін салыстыру. Зерттеу студенттік оқу-зерттеу жұмысы ретінде орындалғандықтан, негізгі назар күрделі өндірістік жүйені құруға емес, әртүрлі модельдердің жұмыс қағидаларын түсіндіруге және олардың нәтижелерін салыстыруға бағытталды.

Тақырып бойынша әдеби шолу. Жаңартылатын энергия генерациясын болжау саласында статистикалық модельдер, машиналық оқыту алгоритмдері және терең оқыту архитектуралары кеңінен қолданылады. Ерте зерттеулерде ARIMA, ARX және басқа авторегрессиялық модельдер қысқа мерзімді күн генерациясын болжау үшін тиімді базалық әдіс ретінде қарастырылды. Мысалы, Bacher және әріптестері күн фотоэлектрлік жүйелерінің 36 сағатқа дейінгі болжамында авторегрессиялық тәсілдердің пайдалы екенін көрсетті (Bacher et al., 2009: 1772).

Күн энергиясын болжауға арналған шолу жұмыстарында болжау дәлдігіне деректер сапасы, уақыттық ажыратымдылық, ауа райы айнымалылары және бағалау метрикалары тікелей әсер ететіні атап өтіледі. Antonanzas және әріптестері фотоэлектрлік қуатты болжауда статистикалық, машиналық оқыту және гибридтік әдістердің қолданылуын жүйеледі (Antonanzas et al., 2016: 78). Voyant және әріптестері машиналық оқыту әдістері күн радиациясын болжауда классикалық модельдерге қарағанда күрделі сызықтық емес тәуелділіктерді жақсы ұстай алатынын көрсетті (Voyant et al., 2017: 569). Sobri және әріптестері күн фотоэлектрлік генерациясын болжауда жасанды нейрондық желілер, тірек векторлық машиналар, регрессиялық модельдер және ансамбльдік тәсілдер жиі қолданылатынын атап өтті (Sobri et al., 2018: 459).

Жел генерациясын болжау күн генерациясына қарағанда күрделірек болуы мүмкін, себебі жел жылдамдығы мен бағыты жоғары стохастикалық сипатқа ие. Жел энергетикасы бойынша шолу еңбектерінде қысқа мерзімді болжам үшін тарихи генерация, метеорологиялық деректер және лагтық белгілерді бірге қолдану ұсынылады (Jung, Broadwater, 2014: 762). Сондықтан жел генерациясын болжауда тек уақыттық қатарға сүйену жеткіліксіз болуы мүмкін, ал сыртқы ауа райы белгілерін қосу модель сапасын арттыра алады.

Машиналық оқыту әдістерінің ішінде шешім ағаштарына негізделген ансамбльдік модельдер практикалық есептерде жиі қолданылады. XGBoost алгоритмі градиенттік бустинг қағидасына негізделеді және кестелік деректердегі сызықтық емес тәуелділіктерді тиімді үйрене алады (Chen, Guestrin, 2016: 785). Энергетикалық болжау есептерінде XGBoost, Random Forest және LightGBM сияқты модельдер лагтық белгілермен, жылжымалы орташа мәндермен және күнтізбелік белгілермен бірге қолданылғанда тұрақты нәтиже береді (Lago et al., 2021: 116983).

Терең оқыту әдістері уақыттық қатарлардағы күрделі тәуелділіктерді үйрену қабілетімен ерекшеленеді. LSTM желісі ұзақ және қысқа мерзімді тәуелділіктерді сақтау үшін арнайы қақпалық механизмдерді қолданады (Hochreiter, Schmidhuber, 1997: 1735). Qing және Niu LSTM моделін күн радиациясын бір күн бұрын болжау міндетінде қолданып, ауа райы болжамдары мен тарихи деректерді бірге пайдалану нәтиженің жақсаруына әсер ететінін көрсетті (Qing, Niu, 2018: 461). GRU, CNN-LSTM және басқа нейрондық архитектуралар да жаңартылатын энергияны болжауда қолданылады, бірақ оларды сапалы оқыту үшін жеткілікті көлемдегі деректер мен есептеу ресурстары қажет.

Соңғы жылдары Transformer негізіндегі модельдер ұзын уақыттық қатарлармен жұмыс істеуге бейімделіп келеді. Informer моделі ұзақ реттіліктерді болжау үшін тиімді attention механизмін ұсынды (Zhou et al., 2021: 11106), ал Temporal Fusion Transformer көпқадамды уақыттық болжамда интерпретацияланатын архитектура ретінде қарастырылады (Lim et al., 2021: 1748). Дегенмен мұндай модельдер студенттік деңгейдегі шағын эксперименттер үшін күрделі болуы мүмкін.

Болжам нәтижелерін бағалау кезінде RMSE, MAE және MAPE метрикалары жиі

қолданылады. Hong және әріптестері энергетикалық болжауда тек бір метрикаға сүйенбей, модель сапасын бірнеше көрсеткіш арқылы бағалау маңызды екенін атап өтті (Hong et al., 2016: 896). Осы әдебиеттерге сүйене отырып, бұл мақалада студенттік зерттеу деңгейінде үш тәсіл салыстырылады: классикалық ARIMA, лагтық белгілерге негізделген XGBoost және LSTM логикасына ұқсас терезелік прокси-модель.

Деректер жиыны. Зерттеуде 2023 күнтізбелік жылына арналған синтетикалық сағаттық деректер жиыны пайдаланылды. Жалпы бақылаулар саны 8760 сағатты құрады. Деректер нақты ENTSO-E деректерінен тікелей алынған жоқ; олар күн және жел генерациясының жалпы статистикалық және маусымдық қасиеттерін имитациялау мақсатында жасалды. Бұл тәсіл студенттік зерттеу үшін модельдерді қауіпсіз және қайталанатын ортада салыстыруға мүмкіндік береді.

Күн генерациясы үшін тәуліктік синусоидалық қисық, маусымдық модуляция және кездейсоқ шу қолданылды. Түнгі уақытта генерация нөлге жақын, ал күндіз жоғары мәндерге ие болады. Жел генерациясы үшін авторегрессивті шу және маусымдық компонент енгізілді, себебі жел ресурсы күн генерациясына қарағанда әлдеқайда стохастикалық сипатқа ие.

Машиналық оқыту модельдері үшін келесі белгілер қолданылды: лагтық белгілер (t-1, t-2, t-3, t-24, t-48, t-168), 24 сағаттық жылжымалы орташа мән және күнтізбелік белгілер (сағат, ай, апта күні). Терезелік прокси-модель үшін соңғы 72 сағаттық деректер терезесі пайдаланылды.

Материалдар мен зерттеу әдістері. ARIMA моделі күн генерациясының күнделікті жиынтық мәндеріне қолданылды. Бұл модель осы жұмыста қарапайым статистикалық базалық тәсіл ретінде қарастырылды. ARIMA нәтижелері сағаттық XGBoost және прокси-модель нәтижелерімен тікелей тең деңгейде емес, базалық салыстыру ретінде ғана талданды.

XGBoost моделі күн және жел генерациясының сағаттық қатарлары үшін бөлек оқытылды. Модель лагтық және күнтізбелік белгілерді пайдаланып, болашақ генерация мәнін болжады. Негізгі гиперпараметрлер ретінде 300 ағаш, максималды тереңдік 6 және оқу жылдамдығы 0,05 қолданылды.

Терең оқытуға ұқсас прокси-модель толық LSTM желісі емес. Ол соңғы 72 сағаттық терезеден алынған сызықтық емес белгілер мен статистикалық сипаттамаларға негізделді. Бұл тәсіл LSTM желісінің уақыттық тәуелділіктермен жұмыс істеу идеясын жеңілдетілген түрде көрсету үшін қолданылды. Мұндай модель оқу-зерттеу мақсатында түсінікті және есептеу жағынан жеңіл болып табылады.

Модельдердің сапасын бағалау үшін RMSE, MAE және MAPE метрикалары қолданылды. Негізгі салыстыру RMSE және MAE бойынша жүргізілді, себебі MAPE нөлге жақын мәндер кезінде тұрақсыз нәтиже беруі мүмкін.

Бағалау хаттамасы. Деректер жиыны уақыт бойынша екі бөлікке бөлінді. 2023 жылғы қаңтар–қараша аралығындағы деректер модельдерді оқыту үшін, ал желтоқсан айының соңғы 30 күні тестілік кезең ретінде пайдаланылды. Мұндай бөлу уақыттық қатарлар үшін дұрыс тәсіл болып саналады, себебі болашақ деректер оқыту кезеңіне араластырылмайды.

Сағаттық модельдер үшін тест кезеңі 720 бақылаудан тұрды. ARIMA моделі күнделікті жиынтық мәндермен жұмыс істегендіктен, оның тест кезеңі 30 күндік бақылаудан тұрды. Осы айырмашылық нәтижелерді түсіндіру кезінде ескерілді.

Нәтижелер. 1-кестеде модельдердің тестілік кезеңдегі нәтижелері көрсетілген. Күн генерациясы бойынша ең төмен RMSE мәні терезелік прокси-модельде байқалды. XGBoost нәтижесі оған өте жақын болды. Жел генерациясында екі модельдің RMSE мәндері шамалас болды, бұл жел генерациясының кездейсоқ ауытқулары жоғары екенін көрсетеді.

Кесте 1. Модельдердің болжау нәтижелері [Table 1 – Model Forecasting Results]

Модель / Model	**RMSE**	**MAE**	**MAPE (%)**
ARIMA (күн/solar, daily)	130,23 MWh	104,60 MWh	11,4
XGBoost (күн/solar)	23,44 MW	18,35 MW	90,0*
XGBoost (жел/wind)	179,92 MW	144,82 MW	105,8*
Прокси-модель/Proxy (күн/solar)	21,50 MW	16,42 MW	72,0*
Прокси-модель/Proxy (жел/wind)	179,71 MW	139,88 MW	113,5*

** Ескерту: MAPE мәндері нөлге жақын генерация кезеңдерінде жоғары болуы мүмкін, сондықтан негізгі салыстыру RMSE және MAE бойынша жүргізілді.*

Күн генерациясы бойынша: прокси-модель XGBoost-тан сәл жақсы нәтиже берді (21,50 vs 23,44 МВт RMSE). ARIMA тәуліктік жиынтық деректермен жұмыс істейді (МВт·сағ), сондықтан оның мәні сағаттық модельдермен тікелей салыстырылмайды — тек базалық бағдар ретінде қарастырылады.

Жел генерациясы бойынша: екі модельдің нәтижесі бірдей деңгейде (179,92 vs 179,71 МВт RMSE). Бұл жел генерациясының жоғары стохастикалық сипатын және тек тарихи деректерге сүйенудің шектеулілігін көрсетеді. Ауа райы айнымалыларын қосу болжам дәлдігін арттыруы мүмкін.

MAE абсолюттік орташа қатені өлшейді. Прокси-модель күн генерациясы үшін ең төмен MAE мәнін берді (16,42 МВт), бұл оның тәуліктік заңдылықтарды XGBoost-қа қарағанда сәл жақсы ұстайтынын көрсетеді.

MAPE нөлге жақын генерация кезеңдерінде (түнгі сағаттар, жел жоқ кездер) жоғары мән береді — бұл модельдің нашарлығын емес, метриканың шектеулілігін көрсетеді. ARIMA-ның 11,4% деп көрінетін нәтижесі оның тәуліктік деректермен жұмыс істеуінен туындайды. Негізгі салыстыру үшін RMSE және MAE қолданылуы ұсынылады.

Талқылау. Күн генерациясы бойынша алынған нәтижелер лагтық және терезелік белгілерді қолданатын модельдердің классикалық статистикалық тәсілдерге қарағанда тиімді болатынын көрсетті. Прокси-модель XGBoost-тан сәл жақсы нәтиже берді, бірақ айырмашылық өте үлкен емес. Бұл студенттік зерттеу үшін маңызды қорытынды: күрделі модель әрқашан айтарлықтай артықшылық бере бермейді, әсіресе деректер көлемі шектеулі болған жағдайда.

Жел генерациясын болжау күрделірек болды. XGBoost пен прокси-модельдің нәтижелері бір-біріне жақын болды. Бұл желдің кездейсоқ сипаты мен сыртқы метеорологиялық факторлардың маңызды екенін көрсетеді. Егер нақты ауа райы болжамдары, жел жылдамдығы, қысым және температура сияқты қосымша белгілер қолданылса, модельдердің сапасы жақсаруы мүмкін.

MAPE метрикасының жоғары мәндері модельдердің міндетті түрде нашар екенін білдірмейді. Күн генерациясы түнде нөлге жақын болады, ал жел генерациясы кейбір кезеңдерде өте төмен болуы мүмкін. Мұндай жағдайда аз абсолюттік қате де үлкен пайыздық қате ретінде көрінеді. Сондықтан жаңартылатын энергияны болжауда RMSE, MAE және нормаланған RMSE сияқты метрикаларды бірге қолдану қажет (Hong et al., 2016: 896).

Зерттеудің шектеулері. Бұл жұмыстың бірнеше шектеуі бар. Біріншіден, зерттеу нақты өндірістік деректерге емес, синтетикалық деректерге негізделді. Сондықтан нәтижелер нақты электр станциялары үшін тікелей қолдануға арналмаған. Екіншіден, толық LSTM нейрондық желісі емес, оның логикасын түсіндіретін жеңілдетілген прокси-модель қолданылды. Үшіншіден, ARIMA моделі күнделікті деректерге, ал XGBoost және прокси-модель сағаттық деректерге қолданылды. Осы себепті ARIMA нәтижелері тек базалық бағдар ретінде қарастырылды.

Болашақ зерттеулерде нақты ENTSO-E, Open Power System Data немесе Renewables.ninja деректерін пайдалану, сондай-ақ толық LSTM, GRU және LightGBM модельдерін қосу ұсынылады. Сонымен қатар ауа райы айнымалыларын енгізу модельдердің практикалық маңызын арттырады.

Қорытынды. Бұл мақалада жаңартылатын энергия генерациясын қысқа мерзімді болжау үшін ARIMA, XGBoost және LSTM логикасына ұқсас терезелік прокси-модель салыстырылды. Зерттеу синтетикалық сағаттық деректер негізінде орындалды және бакалавриат деңгейіндегі оқу-зерттеу жұмысы ретінде қарастырылды.

Нәтижелер күн генерациясы үшін прокси-модельдің ең төмен RMSE мәнін көрсеткенін (21,50 МВт), ал XGBoost моделінің оған жақын нәтиже бергенін (23,44 МВт) анықтады. Жел генерациясы үшін екі модельдің сапасы шамалас болды. Бұл жел генерациясын болжауда сыртқы метеорологиялық факторлардың маңызын көрсетеді.

Жалпы алғанда, жұмыс машиналық оқыту модельдерінің жаңартылатын энергияны болжау міндеттерінде пайдалы екенін көрсетті. Сонымен бірге зерттеу MAPE метрикасын абайлап қолдану қажет екенін және нақты деректермен жұмыс істеу болашақ зерттеулер үшін маңызды бағыт екенін дәлелдеді.

Қосымша: Python кодына қысқаша сипаттама. Зерттеуде Python тіліндегі numpy, pandas, statsmodels, xgboost және scikit-learn кітапханалары қолданылды. Кодтың негізгі кезеңдері: синтетикалық деректерді генерациялау, уақыт бойынша train/test бөлу, лагтық белгілерді құру, XGBoost моделін оқыту, терезелік прокси-модельді құру және RMSE, MAE, MAPE метрикаларын есептеу.

Деректер көздері.

- Renewables.ninja — hourly solar and wind capacity factors, CSV format.
- ENTSO-E Transparency Platform — actual generation data by production type, hourly resolution.
- Open Power System Data — cleaned European power-system time series.
- EIA Open Data — electricity generation data for the United States.

Әдебиеттер

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. Solar Energy, 136, 78–111. https://doi.org/10.1016/j.solener.2016.06.069

Bacher, P., Madsen, H., Nielsen, H. A. (2009). Online short-term solar power forecasting. Solar Energy, 83(10), 1772–1783. https://doi.org/10.1016/j.solener.2009.05.016

Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, T., Pinson, P., Fan, S. (2016). Global Energy Forecasting Competition 2014: Probabilistic energy forecasting. International Journal of Forecasting, 32(3), 896–913. https://doi.org/10.1016/j.ijforecast.2016.02.001

Hyndman, R. J., Athanasopoulos, G. (2021). Forecasting: Principles and Practice. 3rd ed. OTexts.

Jung, J., Broadwater, R. P. (2014). Current status and future advances for wind speed and power forecasting. Renewable and Sustainable Energy Reviews, 31, 762–777. https://doi.org/10.1016/j.rser.2013.12.054

Lago, J., Marcjasz, G., De Schutter, B., Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy, 293, 116983. https://doi.org/10.1016/j.apenergy.2021.116983

Lim, B., Arik, S. O., Loeff, N., Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Pfenninger, S., Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy, 114, 1251–1265. https://doi.org/10.1016/j.energy.2016.08.060

Qing, X., Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy, 148, 461–468. https://doi.org/10.1016/j.energy.2018.01.177

Sobri, S., Koohi-Kamali, S., Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156, 459–497. https://doi.org/10.1016/j.enconman.2017.11.019

Staffell, I., Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. Energy, 114, 1224–1239. https://doi.org/10.1016/j.energy.2016.08.068

Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106–11115.

МРНТИ 28.23.25 DOI: https://doi.org/10.62687/STJ.2.2.2026.24

РЕКОМЕНДАЦИЯ КУЛЬТУРНО-РАЗВЛЕКАТЕЛЬНЫХ СОБЫТИЙ В УСЛОВИЯХ ЭКСТРЕМАЛЬНОГО ХОЛОДНОГО СТАРТА: ЦЕННОСТЬ АФИШИ, ПОГОДНОГО И КАЛЕНДАРНОГО КОНТЕКСТА (НА ДАННЫХ БИЛЕТНОЙ ПЛАТФОРМЫ КАЗАХСТАНА)

[1]Б.Б. Маратов* , **[2]Ж. Ламашева**
[1,2]НАО Евразийский национальный университет имени Л. Н. Гумилёва, г. Астана, Казахстан
*e-mail: 040901550381@enu.kz

Б.Б. Маратов – магистрант 1 курса образовательной программы «Бизнес-аналитика в IT», НАО «Евразийский национальный университет имени Л. Н. Гумилёва», г. Астана, Республика Казахстан, e-mail: maratov.b@enu.kz, https://orcid.org/0009-0001-9053-3649
Ж. Ламашева – PhD, ассоциированный профессор, НАО «Евразийский национальный университет имени Л. Н. Гумилёва», г. Астана, Республика Казахстан.

Аннотация. Настоящая статья посвящена задаче персонализированной рекомендации культурно-развлекательных событий в условиях экстремального холодного старта, характерного для билетных онлайн-платформ. Исследование выполнено на реальных обезличенных данных национальной билетной платформы Казахстана, охватывающих 12 715 взаимодействий 10 657 пользователей с 2 074 событиями в четырёх городах страны за период с марта 2023 г. по февраль 2024 г. Особенностью данных являются медианное число взаимодействий на пользователя, равное единице, доля «холодных» пользователей в контрольном месяце 83 % и около 50 % покупок, приходящихся на события-премьеры, не имеющие истории продаж. В рамках единого протокола строгой временной валидации на шести месячных фолдах сопоставлены 13 алгоритмов: от базовых моделей популярности до гибридного градиентного ранжировщика LightGBM (LambdaRank) с погодными, календарными и афишными признаками и смешиванием рангов методом обратной ранговой фузии (RRF). Статистическая значимость различий оценивалась методом парного бутстрепа (5000 репликатов, 95 % доверительные интервалы). Установлено, что знание афиши следующего месяца обеспечивает наибольший и статистически значимый прирост качества (ΔMAP@10 = +0,0084; $p < 0,0001$), превышающий эффект от любых усложнений собственно алгоритма ранжирования; лучшая конфигурация достигает MAP@10 = 0,0578, что на 17 % выше сильнейшего неперсонализированного базового алгоритма. Погодный контекст демонстрирует прирост на границе статистической значимости ($p = 0,053$), календарные признаки сами по себе значимого эффекта не дают. Полученные результаты обосновывают приоритет интеграции данных афиши в продуктивные рекомендательные контуры событийных платформ.

Ключевые слова: рекомендательные системы, холодный старт, ранжирование, градиентный бустинг, LambdaRank, контекстно-зависимые рекомендации, событийные платформы, MAP@10.

EVENT RECOMMENDATION UNDER EXTREME COLD-START CONDITIONS: THE VALUE OF SCHEDULE, WEATHER AND CALENDAR CONTEXT (BASED ON DATA FROM A KAZAKHSTANI TICKETING PLATFORM)

[1]B.B. Maratov*, [2]Zh. Lamasheva
[1,2]L. N. Gumilyov Eurasian National University, Astana, Kazakhstan
*e-mail: 040901550381@enu.kz

B.B. Maratov – 1st year Master's student of the educational program «Business Analytics in IT», L. N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: maratov.b@enu.kz, https://orcid.org/0009-0001-9053-3649

Zh. Lamasheva – PhD, Associate Professor, L. N. Gumilyov Eurasian National University, Astana, Kazakhstan.

Abstract. This article addresses the problem of personalized recommendation of cultural and entertainment events under the extreme cold-start conditions typical of online ticketing platforms. The study is based on real anonymized data from a national ticketing platform of Kazakhstan, covering 12,715 interactions of 10,657 users with 2,074 events in four cities of the country from March 2023 to February 2024. The dataset is characterized by a median of one interaction per user, an 83 % share of cold-start users in the holdout month, and about 50 % of purchases falling on premiere events with no sales history. Within a unified protocol of strict temporal validation on six monthly folds, 13 algorithms were compared: from basic popularity models to a hybrid gradient-boosted ranker LightGBM (LambdaRank) enriched with weather, calendar and schedule features and rank blending via reciprocal rank fusion (RRF). Statistical significance was assessed using a paired bootstrap procedure (5,000 replicates, 95 % confidence intervals). It was found that knowledge of the next month's event schedule provides the largest and statistically significant quality gain (ΔMAP@10 = +0.0084; $p < 0.0001$), exceeding the effect of any refinement of the ranking algorithm itself; the best configuration achieves MAP@10 = 0.0578, which is 17 % above the strongest non-personalized baseline. The weather context demonstrates a gain at the boundary of statistical significance ($p = 0.053$), while calendar features alone yield no significant effect. The results substantiate the priority of integrating schedule data into production recommendation pipelines of event platforms.

Keywords: recommender systems, cold start, learning to rank, gradient boosting, LambdaRank, context-aware recommendation, event platforms, MAP@10.

ЭКСТРЕМАЛДЫ «СУЫҚ СТАРТ» ЖАҒДАЙЫНДА МӘДЕНИ-ОЙЫН-САУЫҚ ІС-ШАРАЛАРЫН ҰСЫНУ: АФИШАНЫҢ, АУА РАЙЫ МЕН КҮНТІЗБЕЛІК КОНТЕКСТІҢ ҚҰНДЫЛЫҒЫ (ҚАЗАҚСТАНДЫҚ БИЛЕТ ПЛАТФОРМАСЫНЫҢ ДЕРЕКТЕРІ НЕГІЗІНДЕ)

[1]Б.Б. Маратов*, [2]Ж. Ламашева
[1,2]Л. Н. Гумилев атындағы Еуразия ұлттық университеті КеАҚ, Астана қ., Қазақстан
*e-mail: 040901550381@enu.kz

Б.Б. Маратов – «IT-дегі бизнес-аналитика» білім беру бағдарламасының 1-курс магистранты, «Л. Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ, Астана қ., Қазақстан Республикасы, e-mail: maratov.b@enu.kz, https://orcid.org/0009-0001-9053-3649

Ж. Ламашева – PhD, қауымдастырылған профессор, «Л. Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ, Астана қ., Қазақстан Республикасы.

Аңдатпа. Бұл мақала билет сату онлайн-платформаларына тән экстремалды «суық старт» жағдайында мәдени-ойын-сауық іс-шараларын дербестендірілген түрде ұсыну мәселесіне арналған. Зерттеу Қазақстанның ұлттық билет платформасының нақты иесіздендірілген деректері негізінде орындалды: 2023 жылғы наурыз бен 2024 жылғы ақпан аралығында елдің төрт қаласындағы 2 074 іс-шарамен 10 657 пайдаланушының 12 715 өзара әрекеттесуі қамтылды. Деректердің ерекшелігі - бір пайдаланушыға шаққандағы өзара әрекеттесулердің медианасы бірге тең, бақылау айындағы «суық» пайдаланушылардың үлесі 83 % және сатып алулардың шамамен 50 %-ы сату тарихы жоқ премьера-іс-шараларға тиесілі. Қатаң уақыттық валидацияның бірыңғай хаттамасы аясында алты айлық фолдта 13 алгоритм салыстырылды: базалық танымалдылық модельдерінен бастап ауа райы, күнтізбелік және афишалық белгілермен байытылған әрі рангілерді кері рангілік фузия (RRF) әдісімен араластыратын гибридті градиенттік LightGBM (LambdaRank) ранжирлеушісіне дейін. Айырмашылықтардың статистикалық маңыздылығы жұптық бутстреп әдісімен бағаланды (5000 репликат, 95 % сенімділік интервалдары). Келесі айдың афишасын білу сапаның ең

үлкен әрі статистикалық маңызды өсімін қамтамасыз ететіні анықталды (ΔMAP@10 = +0,0084; p < 0,0001), бұл ранжирлеу алгоритмінің өзін кез келген күрделендіру әсерінен асып түседі; ең үздік конфигурация MAP@10 = 0,0578 мәніне жетеді, бұл ең күшті дербестендірілмеген базалық алгоритмнен 17 %-ға жоғары. Ауа райы контексті статистикалық маңыздылық шекарасындағы өсімді көрсетеді (p = 0,053), ал күнтізбелік белгілер өздігінен маңызды әсер бермейді. Алынған нәтижелер афиша деректерін іс-шара платформаларының өндірістік ұсыныс контурларына интеграциялаудың басымдығын негіздейді.

Кілт сөздер: ұсыныс жүйелері, суық старт, ранжирлеу, градиенттік бустинг, LambdaRank, контекстке тәуелді ұсыныстар, іс-шара платформалары, MAP@10.

Введение. Стремительная цифровизация сферы культурного потребления привела к тому, что билетные онлайн-платформы стали основным каналом дистрибуции билетов на кинопоказы, концерты, театральные постановки, спортивные и детские мероприятия. В условиях непрерывно обновляющейся афиши пользователь сталкивается с проблемой информационной перегрузки, а платформа - с необходимостью персонализации выдачи: своевременная и релевантная рекомендация напрямую влияет на конверсию, удержание аудитории и экономику сервиса. Основным инструментом решения этой задачи являются рекомендательные системы, которые за последние два десятилетия стали одним из наиболее востребованных приложений методов машинного обучения (Ricci и др., 2015:1003; Гомзин&Коршунов, 2012:401-417).

Классический арсенал рекомендательных систем включает коллаборативную фильтрацию, основанную на схожести объектов (Sarwar и др., 2001:285-295) или латентных факторных представлениях, получаемых матричной факторизацией (Koren и др., 2009:30-37), в том числе для неявной обратной связи (Hu и др., 2008:263-272); контентные методы, опирающиеся на векторизацию текстовых описаний объектов (Salton&Buckley, 1988:513-523); а также гибридные архитектуры, комбинирующие несколько источников сигнала (Burke, 2002:331-370). Современным стандартом промышленного ранжирования выступают методы обучения ранжированию на основе градиентного бустинга - семейство LambdaRank/LambdaMART (Burges, 2010:19), эффективно реализованное в библиотеке LightGBM (Ke и др., 2017:3146-3154). Для устойчивого объединения нескольких ранжированных списков широко применяется обратная ранговая фузия (reciprocal rank fusion, RRF), демонстрирующая результаты на уровне и выше обучаемых методов смешивания (Cormack и др., 2009:758-759).

Вместе с тем событийный домен принципиально отличается от традиционных доменов рекомендательных систем - кино- и видеосервисов, электронной коммерции, музыкальных платформ. Событие существует ограниченное время и после проведения безвозвратно покидает каталог; жизненный цикл объекта составляет дни или недели; к моменту формирования рекомендации значительная часть релевантных объектов ещё не имеет ни одного взаимодействия (Macedo и др., 2015:123-130). Тем самым проблема холодного старта (Schein и др., 2002:253-260; Lam и др., 2008:208-211) приобретает в событийном домене экстремальный характер: холодными оказываются одновременно и пользователи, и объекты. Кроме того, посещение офлайн-события, в отличие от потребления цифрового контента, потенциально зависит от внешнего контекста - погодных условий, праздничных и выходных дней (Adomavicius&Tuzhilin, 2005:734-749; Trattner и др., 2018:117-150), что делает событийные рекомендации естественным полигоном для контекстно-зависимых моделей.

Анализ литературы выявляет существенный разрыв между академическими бенчмарками и практическими условиями работы региональных билетных платформ. Абсолютное большинство публикуемых сравнений выполняется на плотных наборах данных типа MovieLens или Netflix, где на пользователя приходятся десятки и сотни оценок; при этом показано, что на разреженных данных тщательно настроенные простые базовые алгоритмы нередко превосходят сложные нейросетевые архитектуры (Dacrema и др., 2019:101-109). Систематические исследования на реальных транзакционных данных билетных платформ

Центральной Азии, характеризующихся малым объёмом, экстремальной разреженностью, двуязычным (русским и казахским) описанием контента и выраженной городской локализацией спроса, в открытой печати практически отсутствуют. Не изучен и вопрос о сравнительной ценности различных источников информации - истории продаж, контента, внешнего контекста и афиши предстоящего месяца - в подобных условиях.

Научная новизна настоящего исследования состоит в следующем: (1) впервые на открытых обезличенных данных национальной билетной платформы Республики Казахстан выполнен систематический бенчмарк 13 рекомендательных алгоритмов в рамках единого протокола строгой временной валидации; (2) предложена и количественно оценена концепция «ценности афиши»: показано, что знание каталога событий предстоящего месяца обеспечивает статистически значимый прирост качества, превышающий эффект от усложнения собственно модели ранжирования; (3) впервые для событийного домена Казахстана исследована интеграция погодной климатологии и праздничного календаря в градиентный ранжировщик, построенная без утечек данных из будущего; (4) применена строгая процедура статистической верификации различий на основе парного бутстрепа по наблюдениям «фолд-пользователь», редко используемая в прикладных работах по рекомендательным системам.

Целью работы является количественная оценка сравнительного вклада алгоритмической сложности, контентных сигналов, внешнего контекста и знания афиши в качество рекомендаций культурно-развлекательных событий в условиях экстремального холодного старта. Для достижения цели решались следующие задачи:

1. Сформировать воспроизводимый протокол временной валидации и метрики качества, соответствующие практике событийных платформ.
2. Реализовать и сопоставить базовые, коллаборативные, контентные и гибридные модели рекомендаций с последовательным добавлением погодных, календарных и афишных признаков.
3. Оценить статистическую значимость различий между моделями и выявить признаки, вносящие наибольший вклад в ранжирование.

Материалы и методы исследования. Исследование выполнено на обезличенном наборе транзакционных данных национальной билетной онлайн-платформы Казахстана Ticketon, предоставленном в рамках открытого аналитического конкурса Freedom Data Challenge (Data Science Academy, https://dsacademy.kz/fdc2025). Набор охватывает период с марта 2023 г. по февраль 2024 г. и включает 12 715 взаимодействий 10 657 пользователей с 2 074 событиями в четырёх городах: Алматы, Астане, Шымкенте и Павлодаре. Для 4 597 событий доступны метаданные: категория (кино, театр, концерт и др.), текстовое описание на русском и казахском языках, продолжительность, возрастной рейтинг, признак отечественного производства и год выпуска. Каждое взаимодействие снабжено статусом заказа; при построении моделей статусам присваивались веса: оплаченный заказ - 1,0; отменённый или возвращённый - 0,4; созданный, но не оплаченный - 0,2. Текстовые описания очищались от HTML-разметки, гиперссылок и служебных символов; атрибуты пользователя (город, пол, возраст) рассматривались как статические.

Ключевой особенностью набора является экстремальная разреженность: медианное число взаимодействий на пользователя равно единице, лишь 0,1% пар «пользователь-событие» повторяются, 83% пользователей контрольного месяца не имеют истории покупок (холодные пользователи), а около половины всех покупок приходится на события-премьеры, отсутствовавшие в каталоге ранее. Подобная конфигурация практически исключает эффективное применение классической коллаборативной фильтрации в чистом виде и смещает задачу в область гибридных и контекстно-зависимых методов.

Протокол оценивания построен по принципу строгой временной валидации на шести месячных фолдах с датами отсечения с 1 сентября 2023 г. по 1 февраля 2024 г. Для каждого фолда обучающая история включает только взаимодействия, строго предшествующие дате отсечения, а эталоном (ground truth) служат оплаченные покупки в течение соответствующего календарного месяца. Пул кандидатов формировался из событий,

имевших активность в течение 60 дней до даты отсечения. Качество оценивалось метрикой средней точности MAP@10, вычисляемой по официальному скрипту организаторов конкурса, дополнительно рассчитывались Recall@10 и HitRate@10 с раздельным учётом тёплых и холодных пользователей. Итоговая выборка для статистического анализа содержит 6 309 наблюдений вида «фолд-пользователь».

В сравнение включены следующие группы моделей. Базовые модели популярности: GlobalPop и CityPop ранжируют события по популярности с экспоненциальным затуханием во времени (период полураспада 14 суток), во втором случае - с сегментацией по городу пользователя и откатом к глобальному рейтингу. Коллаборативные модели: ItemKNN строит взвешенную по времени разреженную матрицу «пользователь-событие» и оценивает косинусную близость событий; ALS - матричная факторизация для неявной обратной связи (64 латентных фактора, регуляризация 0,05; 30 итераций, коэффициент уверенности 40) (Hu и др., 2008:263-272). Контентная модель Content TF-IDF векторизует двуязычные описания событий (словарь до 60 000 признаков, униграммы и биграммы) и ранжирует кандидатов по косинусной близости к профилю пользователя, агрегированному из векторов просмотренных событий с временным затуханием (Salton&Buckley, 1988:513-523).

Гибридная модель реализована как двухэтапный конвейер «генерация кандидатов - ранжирование». Объединённый список кандидатов формируется из трёх источников: топ-50 по городской популярности, топ-30 по ItemKNN и топ-30 по контентной близости (суммарно 80-120 кандидатов на пользователя). Ранжирование выполняется градиентным бустингом LightGBM с целевой функцией lambdarank и метрикой MAP (400 деревьев, скорость обучения 0,05; 63 листа, минимум 20 наблюдений в листе, субдискретизация строк и признаков 0,9) (Ke и др., 2017:3146-3154; Burges, 2010:19). Вектор описания пары «пользователь-событие» включает 27 базовых признаков шести групп: сигналы популярности с затуханием (в том числе за окна 3, 7, 14 и 30 суток, тренд и ускорение спроса); характеристики жизненного цикла события (возраст с первого появления, давность последней покупки, число уникальных покупателей); контентные и категориальные признаки (категория, продолжительность, возрастной рейтинг, косинусная близость TF-IDF); коллаборативный сигнал ItemKNN; социально-демографические признаки пользователя (город, пол, возраст, совпадение города); поведенческие агрегаты (аффинность к категории и жанру, число взаимодействий, давность последней активности). Для устранения смещения ранжировщика при обучении применялась комбинированная выборка негативных примеров: к позитивам из пула добавлялось до 40 случайных негативов на пользователя. Финальное предсказание опционально смешивалось с городской популярностью методом обратной ранговой фузии RRF со сглаживающей константой 20 (Cormack и др., 2009:758-759), что стабилизирует выдачу в месяцы с аномальной динамикой спроса.

Контекстные расширения строились с исключением утечек информации из будущего. Погодный блок (8 признаков) использует суточные данные открытого метеоархива Open-Meteo (https://open-meteo.com) по четырём городам; климатическая норма месяца (средняя температура, доля снежных дней, доля дней с неблагоприятной погодой - минимум ниже −15 °C либо осадки не менее 1 мм) рассчитывалась по опорному периоду 2010-2022 гг., заведомо предшествующему интеракционным данным, а фактическая погода привлекалась только для дат, уже наступивших к моменту отсечения. Календарный блок (4 признака) кодирует число праздничных дней целевого месяца, наличие крупных праздников и долю праздничных покупок события. Афишный блок (3 признака) отражает знание каталога предстоящего месяца: индикатор новизны события, признак отечественного фильма и разницу между годом выпуска и текущим годом; добавление этого блока сопровождается расширением пула кандидатов событиями объявленной афиши, не имеющими истории продаж.

Статистическая значимость различий между моделями оценивалась методом парного бутстрепа (Efron&Tibshirani, 1993:436): значения AP@10 ресемплировались по 6 309 наблюдениям «фолд-пользователь» с одинаковыми индексами для сравниваемых моделей (5 000 репликатов, фиксированное зерно генератора), рассчитывались 95% перцентильные

доверительные интервалы разности и двусторонние p-значения. Все эксперименты воспроизводимы: исходный код и полные результаты доступны в открытом репозитории https://github.com/m-bikko/article-freedom-datachallange. Использованные наборы данных опубликованы там же: журнал взаимодействий (data/train_test.csv), метаданные событий (data/events_description.csv) и суточные погодные данные (data/weather_daily.csv).

Результаты и обсуждение. Сводные характеристики использованного набора данных приведены в таблице 1. Обращает на себя внимание сочетание факторов, каждый из которых в отдельности считается серьёзным осложнением для рекомендательных систем, а в совокупности они формируют режим экстремального холодного старта: единичная медиана взаимодействий исключает построение содержательного профиля для большинства пользователей, а преобладание премьер среди покупок означает, что классический коллаборативный сигнал недоступен для значительной части целевых объектов.

Таблица 1. Основные характеристики набора данных билетной платформы (март 2023 г. - февраль 2024 г.)

Показатель	**Значение**
Число взаимодействий	12 715
Число пользователей	10 657
Число событий с взаимодействиями	2 074
Число событий с метаданными	4 597
Города	Алматы, Астана, Шымкент, Павлодар
Медиана взаимодействий на пользователя	1
Доля холодных пользователей в контрольном месяце	83 %
Доля покупок на события-премьеры	≈ 50%
Доля повторных пар «пользователь-событие»	0,1 %

Результаты сравнения моделей по метрике MAP@10, усреднённые по шести месячным фолдам, представлены в таблице 2 и на рисунке 1. Верхней границей достижимого качества служит оракульная модель, ранжирующая только фактически купленные события из пула кандидатов: MAP@10 = 0,4653. Разрыв между оракулом и лучшей практической моделью показывает, что доминирующим ограничением является не алгоритм ранжирования, а принципиальная непредсказуемость единичных покупок при столь разреженной истории.

Таблица 2. Качество моделей рекомендаций по метрике MAP@10 (среднее по шести фолдам; в скобках - 95% бутстреп-интервалы)

Модель	**MAP@10**	**95 % ДИ**
Oracle (потолок пула кандидатов)	0,4653	[0,453; 0,478]
Hybrid + Context (календарь, без RRF)	0,0405	[0,037; 0,044]
GlobalPop (глобальная популярность)	0,0420	[0,039; 0,045]
Hybrid LGBM (чистый ранжировщик)	0,0424	[0,039; 0,046]
Hybrid + Weather (погода, без RRF)	0,0428	[0,039; 0,047]
ALS (матричная факторизация)	0,0465	[0,043; 0,050]
ItemKNN (коллаборативная фильтрация)	0,0479	[0,044; 0,052]

Модель	MAP@10	95 % ДИ
Hybrid + Context + Schedule (афиша, без RRF)	0,0489	[0,045; 0,053]
CityPop (городская популярность)	0,0492	[0,046; 0,053]
Content TF-IDF (контентная модель)	0,0496	[0,046; 0,053]
Hybrid RRF	0,0544	[0,050; 0,059]
Hybrid + Context RRF (календарь)	0,0558	[0,052; 0,060]
Hybrid + Weather RRF (погода)	0,0572	[0,053; 0,062]
Hybrid + Context + Schedule RRF (афиша)	0,0578	[0,054; 0,062]

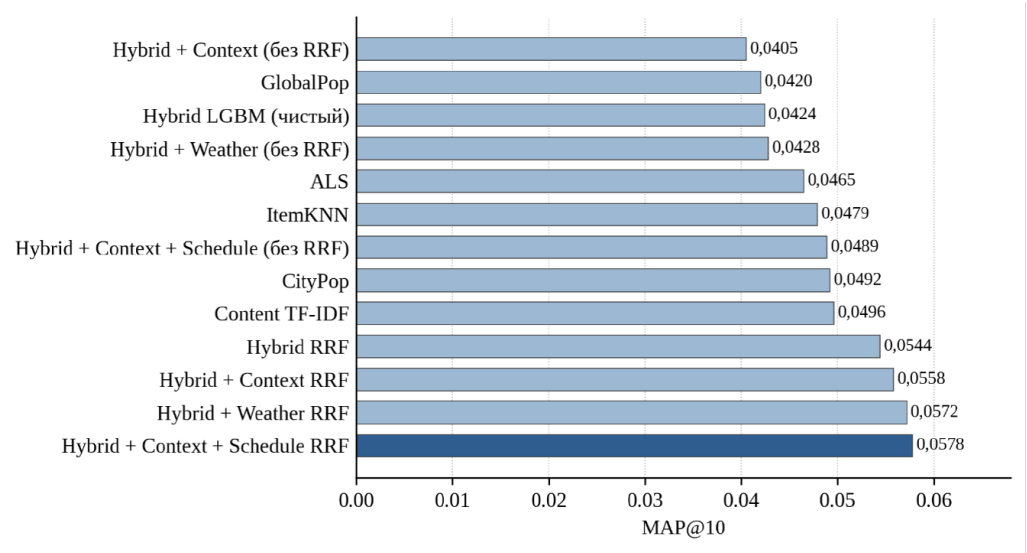


Рисунок 1. Сравнение качества моделей рекомендаций по метрике MAP@10 (среднее по шести месячным фолдам; оракульная модель не показана)

Анализ таблицы 2 позволяет выделить несколько закономерностей. Во-первых, персонализированные коллаборативные модели (ItemKNN, ALS) не превосходят простую городскую популярность: при единичной медиане взаимодействий матрица «пользователь-событие» не содержит достаточного сигнала для обобщения, что согласуется с выводами о силе настроенных базовых алгоритмов на разреженных данных (Dacrema и др., 2019:101-109). Во-вторых, контентная модель на двуязычных TF-IDF представлениях (MAP@10 = 0,0496) оказывается сильнее обоих коллаборативных методов - в условиях дефицита истории текст описания события несёт больше информации, чем следы редких совместных покупок. В-третьих, чистый градиентный ранжировщик уступает базовой городской популярности (0,0424 против 0,0492): при малом числе позитивных примеров помесячная нестабильность обучаемой модели перевешивает её выразительную способность. Смешивание выдачи ранжировщика с популярностным приором методом RRF устраняет этот дефект и выводит гибрид на уровень 0,0544, что статистически значимо выше CityPop ($p = 0,001$). Показательно, что та же закономерность воспроизводится и для контекстных расширений: без RRF конфигурации с погодными (0,0428) и календарными (0,0405) признаками не превосходят базовые алгоритмы, а полная модель с афишей без RRF статистически неотличима от CityPop ($p = 0,896$) - стабилизация популярностным приором является необходимым условием реализации потенциала обучаемого ранжировщика на данных такого объёма.

Результаты проверки статистической значимости ключевых сравнений приведены в таблице 3.

Таблица 3. Парные сравнения моделей (парный бутстреп, 5 000 репликатов, 6 309 наблюдений «фолд-пользователь»)

Сравнение	ΔMAP@10	p-значение	Вывод
Hybrid RRF vs CityPop	+0,0052	0,001	значимо: ранжировщик полезен только в смеси с приором
Hybrid LGBM (чистый) vs CityPop	−0,0069	0,002	значимо хуже: нестабильность по месяцам
+ Афиша vs + Календарь	+0,0084	<0,0001	значимо: наибольший единичный прирост
Лучшая модель vs CityPop	+0,0086	<0,0001	значимо: +17 % к сильнейшему базовому алгоритму
+ Погода vs Hybrid RRF	+0,0028	0,053	на границе значимости
+ Календарь vs + Погода	−0,0022	0,088	значимых различий нет
+ Афиша (без RRF) vs CityPop	−0,0003	0,896	без RRF выигрыш афиши не реализуется

Центральный результат исследования - количественная оценка ценности афиши. Добавление всего трёх признаков, кодирующих знание каталога предстоящего месяца, вместе с расширением пула кандидатов объявленными премьерами даёт прирост ΔMAP@10 = +0,0084 ($p < 0,0001$) - больше, чем любая другая модификация модели, включая переход от простой популярности к полному гибридному конвейеру. Итоговая конфигурация Hybrid + Context + Schedule RRF достигает MAP@10 = 0,0578 [0,054; 0,062], превосходя сильнейший неперсонализированный базовый алгоритм на 17% ($p < 0,0001$). Содержательно это объясняется структурой спроса: поскольку около половины покупок приходится на премьеры, ни одна модель, оперирующая только историческими данными, в принципе не способна рекомендовать значительную часть фактически востребованных событий; информация об афише устраняет это структурное ограничение.

Погодный контекст демонстрирует прирост +0,0028 с p-значением 0,053, формально не достигающим конвенционального порога 0,05: эффект следует квалифицировать как индикативный и требующий проверки на большем объёме данных. Календарные признаки в отрыве от афиши значимого эффекта не дают ($p = 0,088$), что интерпретируется следующим образом: праздничный календарь влияет на общий объём спроса, но слабо дифференцирует события между собой внутри месяца. Анализ важностей признаков ранжировщика (по суммарному приросту качества расщеплений) показывает, что наибольший вклад вносят признаки жизненного цикла события (возраст в каталоге, давность последней покупки), городская популярность и возраст пользователя, а контентная близость TF-IDF систематически превосходит по важности классический коллаборативный сигнал - в полном согласии с результатами прямого сравнения одиночных моделей.

Практическое значение полученных результатов состоит в смещении приоритетов разработки: для событийных платформ, работающих в условиях, аналогичных исследованным, интеграция машиночитаемой афиши предстоящего периода в рекомендательный контур (включая генерацию кандидатов из объявленных, но ещё не продаваемых событий) даёт больший и статистически более надёжный эффект, чем усложнение алгоритма ранжирования, и должна выполняться в первую очередь. Дополнительным практическим выводом является необходимость стабилизации обучаемых ранжировщиков популярностным приором (например, методом RRF) при малых объёмах обучающих данных.

Заключение. В работе впервые на открытых обезличенных данных национальной билетной платформы Казахстана выполнено систематическое сравнение 13 рекомендательных алгоритмов в условиях экстремального холодного старта: при медиане в одно взаимодействие на пользователя, 83% холодных пользователей и около 50% покупок, приходящихся на события

без истории продаж. Единый протокол строгой временной валидации на шести месячных фолдах в сочетании с парным бутстрепом по 6 309 наблюдениям обеспечил статистическую корректность выводов.

Главный вывод исследования формулируется как принцип «ценности афиши»: знание каталога событий предстоящего месяца обеспечивает наибольший статистически значимый прирост качества рекомендаций (ΔMAP@10 = +0,0084; $p < 0,0001$), превышающий эффект любых усложнений собственно модели. Лучшая конфигурация - гибридный градиентный ранжировщик LightGBM с погодными, календарными и афишными признаками, стабилизированный обратной ранговой фузией с городской популярностью, - достигает MAP@10 = 0,0578, что на 17% выше сильнейшего базового алгоритма. Установлено также, что в условиях экстремальной разреженности контентные сигналы двуязычных описаний превосходят классическую коллаборативную фильтрацию, чистый обучаемый ранжировщик без популярностного приора неустойчив, а календарные признаки сами по себе значимого эффекта не дают; влияние погодного контекста находится на границе статистической значимости и требует дальнейшей проверки.

Ограничениями работы являются охват одного годового цикла и четырёх городов, оценка исключительно в офлайн-режиме без онлайн-экспериментов, а также использование фиксированного набора табличных признаков. Перспективными направлениями дальнейших исследований представляются: применение последовательностных и нейросетевых моделей (Quadrana и др., 2018:66) и многоязычных векторных представлений описаний событий; расширение погодного и социокультурного контекста; проведение онлайн-A/B-экспериментов для проверки переносимости офлайн-выводов; а также воспроизведение методики на данных других региональных платформ для оценки обобщаемости принципа «ценности афиши».

Литература

Гомзин&Коршунов, 2012 - Гомзин А.Г., Коршунов А.В. Системы рекомендаций: обзор современных подходов // Труды Института системного программирования РАН. - 2012. - Т. 22. - С. 401-417. [Rus]

Adomavicius&Tuzhilin, 2005 - Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering. - 2005. - Vol. 17, № 6. - P. 734-749. DOI: 10.1109/TKDE.2005.99 [Eng]

Burges, 2010 - Burges C.J.C. From RankNet to LambdaRank to LambdaMART: An overview // Microsoft Research Technical Report MSR-TR-2010-82. - 2010. - 19 p. [Eng]

Burke, 2002 - Burke R. Hybrid recommender systems: Survey and experiments // User Modeling and User-Adapted Interaction. - 2002. - Vol. 12, № 4. - P. 331-370. DOI: 10.1023/A:1021240730564 [Eng]

Cormack и др., 2009 - Cormack G.V., Clarke C.L.A., Buettcher S. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods // Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. - 2009. - P. 758-759. DOI: 10.1145/1571941.1572114 [Eng]

Dacrema и др., 2019 - Dacrema M.F., Cremonesi P., Jannach D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches // Proceedings of the 13th ACM Conference on Recommender Systems. - 2019. - P. 101-109. DOI: 10.1145/3298689.3347058 [Eng]

Efron&Tibshirani, 1993 - Efron B., Tibshirani R.J. An Introduction to the Bootstrap. - New York: Chapman & Hall, 1993. - 436 p. [Eng]

Hu и др., 2008 - Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets // Proceedings of the 8th IEEE International Conference on Data Mining. - 2008. - P. 263-272. DOI: 10.1109/ICDM.2008.22 [Eng]

Ke и др., 2017 - Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A highly efficient gradient boosting decision tree // Advances in Neural Information Processing Systems. - 2017. - Vol. 30. - P. 3146-3154. [Eng]

Koren и др., 2009 - Koren Y., Bell R., Volinsky C. Matrix factorization techniques for recommender systems // Computer. - 2009. - Vol. 42, № 8. - P. 30-37. DOI: 10.1109/MC.2009.263 [Eng]

Lam и др., 2008 - Lam X.N., Vu T., Le T.D., Duong A.D. Addressing cold-start problem in recommendation systems // Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication. - 2008. - P. 208-211. DOI: 10.1145/1352793.1352837 [Eng]

Macedo и др., 2015 - Macedo A.Q., Marinho L.B., Santos R.L.T. Context-aware event recommendation in event-based social networks // Proceedings of the 9th ACM Conference on Recommender Systems. - 2015. - P. 123-130. DOI: 10.1145/2792838.2800187 [Eng]

Quadrana и др., 2018 - Quadrana M., Cremonesi P., Jannach D. Sequence-aware recommender systems // ACM Computing Surveys. - 2018. - Vol. 51, № 4. - Article 66. - P. 1-36. DOI: 10.1145/3190616 [Eng]

Ricci и др., 2015 - Ricci F., Rokach L., Shapira B. (eds.) Recommender Systems Handbook. - 2nd ed. - New York: Springer, 2015. - 1003 p. DOI: 10.1007/978-1-4899-7637-6 [Eng]

Salton&Buckley, 1988 - Salton G., Buckley C. Term-weighting approaches in automatic text retrieval // Information Processing & Management. - 1988. - Vol. 24, № 5. - P. 513-523. DOI: 10.1016/0306-4573(88)90021-0 [Eng]

Sarwar и др., 2001 - Sarwar B., Karypis G., Konstan J., Riedl J. Item-based collaborative filtering recommendation algorithms // Proceedings of the 10th International Conference on World Wide Web. - 2001. - P. 285-295. DOI: 10.1145/371920.372071 [Eng]

Schein и др., 2002 - Schein A.I., Popescul A., Ungar L.H., Pennock D.M. Methods and metrics for cold-start recommendations // Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. - 2002. - P. 253-260. DOI: 10.1145/564376.564421 [Eng]

Trattner и др., 2018 - Trattner C., Oberegger A., Marinho L., Parra D. Investigating the utility of the weather context for point of interest recommendations // Information Technology & Tourism. - 2018. - Vol. 19. - P. 117-150. DOI: 10.1007/s40558-017-0100-9 [Eng]

References

Gomzin&Korshunov, 2012 - Gomzin, A. G., & Korshunov, A. V. (2012). Sistemy rekomendatsiy: obzor sovremennykh podkhodov [Recommendation systems: a survey of modern approaches]. Trudy Instituta sistemnogo programmirovaniya RAN, 22, 401-417. [In Russ]
Adomavicius&Tuzhilin, 2005 - Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749. [In Eng]
Burges, 2010 - Burges, C. J. C. (2010). From RankNet to LambdaRank to LambdaMART: An overview (Microsoft Research Technical Report MSR-TR-2010-82). 19 p. [In Eng]
Burke, 2002 - Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370. [In Eng]
Cormack et al., 2009 - Cormack, G. V., Clarke, C. L. A., & Buettcher, S. (2009). Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In Proceedings of the 32nd International ACM SIGIR Conference (pp. 758-759). [In Eng]
Dacrema et al., 2019 - Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 101-109). [In Eng]
Efron&Tibshirani, 1993 - Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall. 436 p. [In Eng]
Hu et al., 2008 - Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 263-272). [In Eng]
Ke et al., 2017 - Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146-3154. [In Eng]
Koren et al., 2009 - Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37. [In Eng]
Lam et al., 2008 - Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (pp. 208-211). [In Eng]
Macedo et al., 2015 - Macedo, A. Q., Marinho, L. B., & Santos, R. L. T. (2015). Context-aware event recommendation in event-based social networks. In Proceedings of the 9th ACM Conference on Recommender Systems (pp. 123-130). [In Eng]
Quadrana et al., 2018 - Quadrana, M., Cremonesi, P., & Jannach, D. (2018). Sequence-aware recommender systems. ACM Computing Surveys, 51(4), Article 66, 1-36. [In Eng]
Ricci et al., 2015 - Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). Recommender Systems Handbook (2nd ed.). New York: Springer. 1003 p. [In Eng]
Salton&Buckley, 1988 - Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513-523. [In Eng]
Sarwar et al., 2001 - Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295). [In Eng]
Schein et al., 2002 - Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In Proceedings of the 25th Annual International ACM SIGIR Conference (pp. 253-260). [In Eng]
Trattner et al., 2018 - Trattner, C., Oberegger, A., Marinho, L., & Parra, D. (2018). Investigating the utility of the weather context for point of interest recommendations. Information Technology & Tourism, 19, 117-150. [In Eng]

ҒТАМР 28.23.00 DOI: https://doi.org/10.62687/STJ.2.2.2026.26

STEM ЖӘНЕ КОМПЬЮТЕРЛІК ҒЫЛЫМДАРДАҒЫ СТУДЕНТТЕРДІҢ ҮЛГЕРІМІН БОЛЖАУДЫ ИНТЕРПРЕТАЦИЯЛАУҒА АРНАЛҒАН ТҮСІНДІРМЕЛІ ЖАСАНДЫ ИНТЕЛЛЕКТ (ХАІ): ЖҮЙЕЛІ ӘДЕБИЕТТІК ШОЛУ

[1]А.Ә. Әлібай*, **[2]Г.К. Тағанова**
[1,2]Астана халықаралық университеті, Астана қ., Қазақстан Республикасы
*e-mail: aidanaalibai9@gmail.com

А.Ә. Әлібай – ВШИТИИ Бакалавриант студенті, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222
Г.К. Тағанова – аға оқытушы, ғылыми жетекші, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аңдатпа. Білім беру деректерін интеллектуалды талдау (Educational Data Mining, EDM) оқушылардың үлгерімін болжау арқылы оқудағы қиындықтарды ерте анықтауға мүмкіндік береді. Дегенмен, Машиналық оқыту үлгілері көбінесе "қара жәшік" принципі бойынша жұмыс істейді. Түсіндірілетін жасанды интеллект (Explainable Artificial Intelligence, XAI) неліктен "қара жәшік" модельдері белгілі бір болжамдарды беретінін түсінуге көмектеседі. Бұл мақалада информатика және жаратылыстану ғылымдары бойынша оқушылардың үлгерімін болжау үшін түсіндірмелі жасанды интеллектті қолдану саласындағы соңғы бес жылдағы зерттеулерге жүйелі шолу берілген. Біз ең жиі қолданылатын сипаттамалар оқушылардың мінез-құлқы мен үлгерімі туралы деректер екенін және болжаудың негізгі мақсаттары пән бойынша үлгермеу қаупімен немесе бағалаулармен байланысты екенін анықтадық. Зерттеу сонымен қатар XAI қолданбаларын қарастырды, нәтижесінде ең көп таралған қолданбалар белгілердің маңыздылығын жаһандық талдау, жеке болжамдарды түсіндіру және араласулар мен шешім қабылдауды қолдау болып табылады. Сонымен қатар, біз Шеплидің аддитивті түсіндірмелері (SHAP) жеке деңгейде шектеулі қолданумен, негізінен жаһандық деңгейде ең көп қолданылатын XAI әдісі екенін анықтадық. Сонымен қатар, курстарды жақсарту, визуализация параметрлері және жекелендірілген ұсыныстарды қалыптастыру үшін түсіндірілетін жасанды интеллектті қолдануға қатысты зерттеулердегі олқылық анықталды. Бұл олқылықты жою мұғалімдерге жеке оқушыларға жақсы қолдау көрсету үшін деректерге негізделген жеке ұсыныстар беруге мүмкіндік береді.

Түйінді сөздер: Түсіндірілетін Жасанды Интеллект, XAI, Білім Беру Деректерін Өндіру, EDM, Өнімділікті Болжау, Информатика Білімі, STEM Білімі, Жүйелі Әдебиеттерге Шолу.

ОБЪЯСНИМЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ (XAI) ДЛЯ ИНТЕРПРЕТАЦИИ ПРОГНОЗОВ УСПЕВАЕМОСТИ СТУДЕНТОВ В ОБЛАСТИ STEM И КОМПЬЮТЕРНЫХ НАУК: СИСТЕМАТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ

[1]А.А. Алибай*, **[2]Г.К. Таганова**
[1,2]Международный университет Астана, г. Астана, Республика Казахстан
*e-mail: aidanaalibai9@gmail.com

А.А. Алибай – студент бакалавриата ВШИТИИ, Международный университет Астана, г. Астана, Казахстан, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222
Г.К. Таганова – старший преподаватель, научный руководитель, Международный университет Астана, г. Астана, Казахстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аннотация. Интеллектуальный анализ образовательных данных (Educational Data Mining, EDM) позволяет выявлять трудности в обучении на ранних этапах, прогнозируя успеваемость учащихся. Однако модели машинного обучения часто работают по принципу «черного ящика». Объяснимый искусственный интеллект (Explainable Artificial Intelligence, XAI) помогает понять, почему модели «черного ящика» выдают те или иные прогнозы. В этой статье представлен систематический обзор исследований, проведенных за последние пять лет в области применения объяснимого искусственного интеллекта для прогнозирования успеваемости учащихся в сфере компьютерных наук и естественно-научного образования. Мы выяснили, что наиболее часто используемыми характеристиками являются данные о поведении и успеваемости учащихся, а основные цели прогнозирования связаны с риском неуспеваемости по предмету или с оценками. В исследовании также рассматривались области применения XAI, в результате чего выяснилось, что наиболее распространенными областями применения являются глобальный анализ важности признаков, объяснение индивидуальных прогнозов и поддержка вмешательств и принятия решений. Более того, мы обнаружили, что аддитивные объяснения Шепли (SHAP) были наиболее часто используемой методикой XAI, применяемой преимущественно на глобальном уровне, с ограниченным использованием на индивидуальном уровне. Кроме того, был выявлен пробел в исследованиях, связанных с использованием объяснимого искусственного интеллекта для улучшения курсов, настройки визуализаций и формирования персонализированных рекомендаций. Устранение этого пробела позволит преподавателям предоставлять персонализированные рекомендации на основе данных, чтобы лучше поддерживать отдельных учащихся.

Ключевые слова: Объяснимый искусственный интеллект, XAI, Интеллектуальный анализ образовательных данных, EDM, прогнозирование производительности, Образование в области компьютерных наук, STEM-образование, Систематический обзор литературы.

EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) FOR INTERPRETING STUDENT PERFORMANCE PREDICTION IN STEM AND COMPUTER SCIENCE: A SYSTEMATIC LITERATURE REVIEW

[1]A.A. Alibai*, **[2]G.K. Taganova**
[1,2]Astana International University, Astana, Republic of Kazakhstan
*e-mail: aidanaalibai9@gmail.com

A.A. Alibai – Bachelor student of the Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222

G.K. Taganova – Senior Lecturer, Research Supervisor, Astana International University, Astana, Kazakhstan, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Annotation. Educational Data Mining (EDM) enables the early identification of learning difficulties through the prediction of students' academic performance. However, machine learning models often operate as "black boxes," making their decision-making processes difficult to interpret. Explainable Artificial Intelligence (XAI) addresses this challenge by providing insights into why these models generate specific predictions. This article presents a systematic literature review of research published over the past five years concerning the application of Explainable Artificial Intelligence (XAI) to predict student performance in computer science and STEM education. The findings indicate that the most frequently used features were students' behavioral data and academic records, while the primary prediction tasks focused on grade prediction and identifying students at risk of academic failure. The study also analyzed the application domains of XAI, identifying global feature importance analysis, the explanation of individual predictions, and support for pedagogical interventions as the most prevalent areas. Furthermore, SHapley Additive exPlanations (SHAP) emerged as the most

widely used XAI method, applied predominantly for global model interpretation, with limited use for explaining individual predictions. Finally, a significant research gap was identified regarding the use of Explainable Artificial Intelligence (XAI) for course optimization, customized visualizations, and the generation of personalized recommendations. Addressing these gaps could enable educators to provide data-driven, personalized support to improve individual student outcomes.

Keywords: Explainable Artificial Intelligence, XAI, Educational Data Mining, EDM, Performance Prediction, Computer Science Education, STEM Education, Systematic Literature Review.

Кіріспе. 21 ғасырда информатика мен STEM-ге қатысты дағдылар маңызды бола бастады. Алайда, бұл пәндер көбінесе оқушыларға қиындық туғызады, өйткені олар математикалық амалдар мен абстрактілі ұғымдардың жиынтығын, бағдарламалау синтаксисін және тіпті тілді түсінуді қажет етеді, бұл жаңадан бастаушыларға қиындық тудырады (Келлехер, Пауш, 2005: 83–137, Чой және т.б., 2024: 1–8).

Білім беру деректерін өндіру (EDM) - машиналық оқытуды, терең оқытуды немесе статистикалық әдістерді пайдалана отырып, білім беру деректерін талдайтын деректерді өндіру бөлімі. EDM қолданудың бірі-оқушылардың үлгерімін болжау (Чой және т.б., 2023). Бұл оқушылардың табысқа жету жолында көмектесетін педагогикалық араласуларға мүмкіндік беру арқылы (Ромеро және т.б., 2008: 368–384) оқу проблемаларын ерте анықтауға ықпал етеді (Бейкер, Йасеф, 2009: 3–17). Дегенмен, машиналық және терең оқытудың дәстүрлі үлгілері әдетте "қара жәшіктер" болып табылады, бұл олардың неліктен нақты болжамдар беретінін түсінуді қиындатады.

Түсіндірілетін жасанды интеллект (XAI) әдістері мұғалімдерге оқушылардың жетістігіне немесе сәтсіздігіне ықпал ететін факторларды түсінуге көмектесу арқылы оқушылардың үлгерімін болжау үлгілерін түсіндіреді (Чой және т.б., 2024ә) (Чой және т.б., 2024a: 1–8). XAI әдістері жасанды интеллект модельдерінің шешімдерін неғұрлым түсіндіруге мүмкіндік береді, модельдердің ашықтығын арттырады және шешім қабылдау процестерін жақсартады (Гилпин және т.б., 2018: 80–89) (Адади, Беррада, 2018: 52138–52160).

Бұл жүйелі әдебиеттерге шолу студенттердің информатика және STEM білім беру саласындағы үлгерімі туралы болжамдарды түсіндірудің заманауи XAI әдістерін қарастырды. Білім берудің осы салаларына ерекше назар аударуымыздың себебі олардың болашақ технологиялық дамуға айтарлықтай әсері бар болып табылады. Зерттеу барысында келесі сұрақтар қойылған болатын.

Осы зерттеуге негізделген сұрақтар:

1. XAI зерттеулерінде оқушылардың үлгерімін түсіндіру үшін қандай мәліметтер жиынтығы, белгілер түрлері, болжау мақсаттары мен алгоритмдер қолданылды?
2. бұл зерттеулерде қандай XAI әдістері, интерпретация деңгейлері, бейнелеу әдістері және практикалық іске асыру қолданылды?

Шолуда әдебиеттегі ықтимал олқылықтарды анықтау және өнімділікті болжау модельдерінің интерпретациясы мен пайдалылығын жақсарту бағыттарын ұсыну үшін осы мәселелер қарастырылды.

Зерттеу материалдары мен әдістері. Қарастырылған әдебиеттерде зерттеушілер әдетте XAI-де "интерпретация" және "түсініктілік" терминдерін бір-бірінің орнына қолданады, бұл адамның жасанды интеллект моделінің неліктен нақты болжам жасағанын түсіну дәрежесін білдіреді (Гилпин және т.б., 2018: 80–89) (Матрани және т.б., 2021: 100060). Алайда, бірнеше зерттеулерде интерпретация мен түсініктілік жеке ұғымдар ретінде қарастырылды (Марцинкевичс, Фогт, 2020), онда интерпретация модельге тән мөлдірлікті білдіреді.

XAI саласындағы алғашқы шолулар интерпретацияның негізгі тұжырымдамаларын анықтауға және XAI әдістерін жіктеуге бағытталған. Адади мен Беррада (Адади, Беррада, 2018: 52138–52160)». 2018 жылы 2004 жылдан 2018 жылға дейін жарияланған 381 мақаланы қамтитын әдебиетке айтарлықтай шолу жасады. Олар XAI әдістерін Машиналық оқыту алгоритмдеріне, интерпретация деңгейіне (жергілікті немесе ғаламдық) және интерпретация уақытына (кірістірілген немесе пост-арнайы) қолданылуына қарай жіктеді.Ішкі интерпретация

олардың арқасында жанама түсіндірілетін модельдерге жатады сызықтық регрессия сияқты салыстырмалы түрде қарапайым құрылым. Фактіден кейінгі интерпретация күрделі "қара жәшік"модельдерінде қабылданған шешімдерді түсіндіру үшін қосымша әдістерді қолдануды қамтиды.

Гидотти және т.б. (Гвидотти және т.б., 2018: 1–42) 2019 жылы XAI әдістерін "қара жәшік" модельдерін түсіндіру, модельдерді тексеру, олардың нәтижелерін түсіндіру және мөлдір модельдерді құру үшін қолданылатын әдістерге бөлу үшін шолу жасады. Олар интерпретацияны бағалаудың ресми көрсеткіштерінің жоқтығын және белгісіз немесе жасырын ерекшеліктерге негізделген модельдерді түсіндірудің қиындығын атап өтті.

Арриета және т.б. (Арриета және т.б., 2020: 82–115) 2020 жылы мөлдір және post-hoc әдістері арасындағы айырмашылыққа және терең оқытуды түсіндіру әдістері үшін нақты Ішкі санаттарды анықтауға бағытталған шолуды ұсынды.

2023 жылы Чауши және оның авторлары (Чауши және т.б., 2023: 48–71) XAI-ге білім беруде шолу жасап, сенімділікті арттыру үшін ашықтық қажеттілігін атап өтті. Олар жасанды интеллекттің күрделі модельдерімен және интерпретацияға қойылатын талаптармен байланысты мәселелерді анықтады және білім берудегі XAI-ге қатысты болашақ зерттеулерге арналған бағыттарды ұсынды.

Алайда, қолданыстағы шолуларда XAI туралы құнды ақпарат болғанымен, олар негізінен XAI-ді білім берудің нақты салаларында қолдануға емес, технологиялық жағына қатысты болды.

Біздің әдебиеттерге жүйелі шолу информатика және STEM білім беру үшін EDM-де XAI әдістерін қолдану бойынша соңғы бес жылдағы (2020-2024) соңғы зерттеулерді талдау арқылы бұл олқылықты жоюға мүмкіндік берді. Біз ақпарат беруге, зерттеулердегі олқылықтарды анықтауға және болашаққа бағыттар ұсынуға тырыстық.

Бұл әдебиетке жүйелі шолу Китченхэм тұжырымдамасына сәйкес (Китченхем, Чартерс, 2007) инженерлік немесе жаратылыстану ғылымдары бағыты бойынша білім беру саласында жүйелі шолулар жүргізілді.

Іздеу стратегиясы. Біз информатика және инженерлік зерттеулер, ACM Digital Library, IEEE Xplore және Web of Science бойынша ең танымал дерекқорларды пайдалана отырып, әдебиеттерді жан-жақты іздедік.

Іздеу жолағы төрт негізгі тұжырымдамаға негізделген: білім беру деректерін өндіру, XAI әдістері, өнімділікті болжау және бағдарламалауды оқыту немесе STEM. Нақтылаудың бірнеше қайталануынан кейін біз іздеу жолағын тұжырымдадық:

("жасанды интеллект" немесе "AI" немесе "машиналық оқыту" немесе "терең оқыту" немесе "білім беру деректерін іздеу" немесе"EDM") және ("SHAP*" немесе "LIME"немесе" модельге тәуелді емес жергілікті түсіндірмелер", Немесе "PDP" немесе "ішінара тәуелділік графигі" немесе "Якорь "немесе" мұз "немесе" жеке шартты күту "немесе" але "немесе" жинақталған жергілікті әсер") және ("түсіндіру* "немесе" түсіндіру* "немесе" транспарациялау*") және ("өнімділік") және ("болжау*") Және ("бағдарламалау*" немесе "CS1" немесе "компьютер" немесе "STEM") және ("білім "немесе" оқыту* "немесе" мектеп "немесе " оқушы" немесе "сынып *" немесе "академиялық" немесе "курс" немесе "оқу бағдарламасы").

Қосу Және Алып Тастау Критерийлері. Біз таңдалған мақалалардың зерттеу міндеттерімізге сәйкес келуін қамтамасыз ету үшін қосу және алып тастау критерийлерін белгіледік.

Мақаланы қосу үшін мынадай критерийлер айқындалды: i) Машиналық оқыту немесе терең оқыту тәсілдерін пайдалана отырып, Информатика немесе STEM-білім беру саласындағы оқушылардың үлгерімін болжайтын зерттеулер; ii) болжау модельдерін түсіндіру үшін XAI әдістері анық қолданылатын зерттеулер; iii) XAI іске асыру әдістерінің нақты сипаттамаларын қамтитын зерттеулер; IV) рецензияланатын журналдарда немесе конференцияларда жарияланған зерттеулер.

Алып тастау критерийлері: i) студенттердің үлгерімін болжаусыз мінез-құлқын талдауға

бағытталған зерттеулер; ii) информатика немесе STEM білімімен байланысты емес зерттеулер; iii) ағылшын тілінде емес мақалалар; iv) эмпирикалық нәтижелерсіз аналитикалық мақалалар, шолулар немесе теориялық жұмыстар.

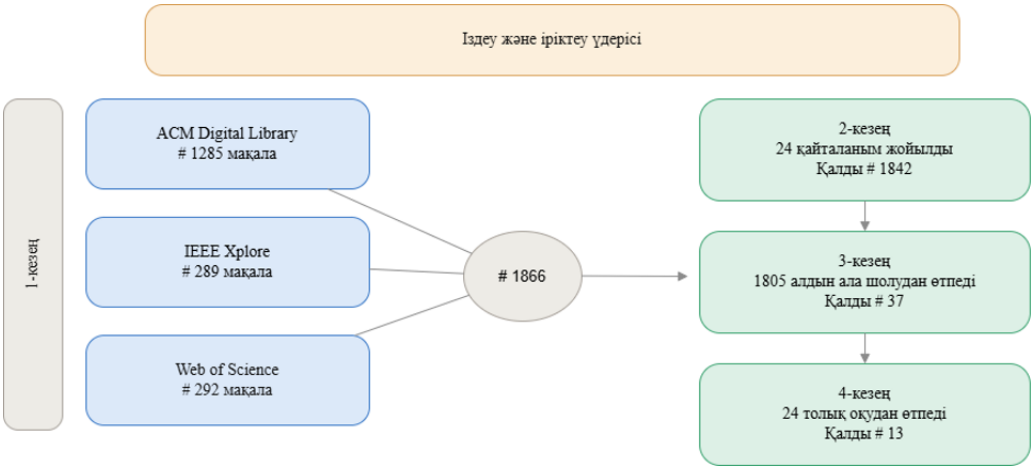
Іздеу және таңдау процесі. Суретте көрсетілгендей, біз іздеу және іріктеу процесінің төрт кезеңін қолданып, кейінгі талдау үшін ең маңызды әдебиеттерді анықтадық.

1 кезең: 2024 жылдың шілдесінде біз дерекқордағы тезистер мен толық мәтіндерге алдын ала анықталған іздеу жолын қолдандық, нәтижесінде 1866 мақала пайда болды.

2 кезең: тақырыпқа, авторларға және DOI-ге негізделген 24 телнұсқаны Алып тастау 1842 мақаланы қалдырды.

3 кезең: біз қосу және алып тастау критерийлерін тексеру үшін тақырыптарды, тезистерді және кілт сөздерді қолмен қарап шықтық, содан кейін кілт сөз тіркестері бойынша мәтіндік іздеу жүргіздік. Бұл тиісті абзацтар мақалалардағы XAI рөлін анықтау үшін оқылды. Бұл қадам бізге XAI-ді тек байланысты жұмыс ретінде атап өткен, бірақ оны оқушылардың үлгерімін болжау үшін арнайы қарастырмаған зерттеулерді алып тастауға мүмкіндік берді. Содан кейін біз 1805 зерттеуді алып тастадық, нәтижесінде 37 зерттеу қалды.

4-қадам: Қалған мақалалардың толық мәтінін іріктеу нәтижесінде егжей-тегжейлі метадеректер (мысалы, атауы, мақсаттары, қорытындылары, білім беру контексті) алынып, өзектілігіне негізделген қосымша алып тастаулар жасалды. Бұл соңғы қадам талдау үшін 13 мақала қалдырды.



Сурет 1. Іздеу және таңдау процесі [Search and selection process]

Зерттеулердің сапасын бағалау (Quality Assessment). Таңдалған зерттеулердің сапасы Kitchenham ұсынған жүйелі әдеби шолуларды жүргізу жөніндегі әдістемелік ұсынымдарға сәйкес бағаланды. Әрбір мақала зерттеу мақсаттарының айқындығы, қолданылған машиналық оқыту алгоритмдерінің негізділігі, Explainable Artificial Intelligence (XAI) әдістерінің сипатталу толықтығы және алынған нәтижелердің негізділігі сияқты критерийлер бойынша сарапталды. Осы талаптарға сәйкес келетін және STEM білімі саласында практикалық маңызы жоғары зерттеулер ғана қорытынды талдауға енгізілді.

Жүйелі қателіктер қаупін талдау (Risk of Bias). Шолу нәтижелерінің объективтілігін қамтамасыз ету мақсатында ықтимал жүйелі қателіктердің көздері талданды. Біріншіден, тек ағылшын тіліндегі жарияланымдардың енгізілуіне байланысты тілдік ауытқу (language bias) болуы мүмкін. Екіншіден, мақалаларды іріктеу қолмен жүргізілгендіктен, таңдау ауытқуының (selection bias) ықтималдығын азайту үшін алдын ала анықталған қосу және алып тастау критерийлері қатаң сақталды. Сонымен қатар, тек рецензияланатын ғылыми жарияланымдарды қамту арқылы жарияланымдық ауытқу (publication bias) ескерілді.

Нәтижелер мен талқылау. Rq1: деректер жиынтығы, сипаттамалары, болжау мақсаты және алгоритмдер. Деректер жиынына келетін болсақ, 1-кестеде көрсетілгендей, зерттеулердің 69,2% - ы өздігінен жиналған деректер жиынын пайдаланды, ал 30,8% -

ы STEM-мен байланысты Open University Learning Analytics (OULAD) деректер жиынын пайдаланды (Кузилек және т.б., 2017: 1–8).

Кесте 1. Деректер жиынтығы мен деректер көздерін тарату [Distribution of datasets and data sources]

Дерек көзі	Қолжетімділік	N	Пайыз
LMS	Өзі жинаған	4	30,8%
LMS	Жалпыға қолжетімді	4	30,8%
AJS	Өзі жинаған	2	15,4%
SIS	Өзі жинаған	2	15,4%
Басқалар	Өзі жинаған	1	7,7%

Деректер көздеріне келетін болсақ, ең көп таралған деректер көзі зерттеудің 61,5%-ында қолданылған оқуды басқару жүйелері (LMS) болды, одан кейін автоматтандырылған бағалау жүйелері (AJS) және оқушылардың ақпараттық жүйелері (SIS) әрқайсысы 15,4%-дан, ал 7,7%-ы оқытушылардың сауалнамалары мен бақылауларына сүйенді (Джаннакас және т.б., 2021).

Іріктеме өлшемдері 100-ден кем емес және 30 000-нан астам оқушыға дейін болды, OULAD сияқты үлкен деректер жиынтығы жоғары болжам дәлдігін қамтамасыз етті. Мысалы Аднан және басқалары (Аднан және т.б., 2022: 129843–129864) жүргізген зерттеу бойынша 91,9% құрады. Кішігірім деректер жиынтығын да құнды ақпарат берді. Мысалы, Рико-Хуан және т.б. (Рико-Хуан және т.б., 2023: 955–969) 90 оқушыдан алынған деректерді пайдалана отырып, AUC 0,70-ке жетті.

Нысан түрлері және кіріс айнымалылары. Болжалды модельдер кіріс ретінде қызмет ететін әртүрлі нысандар негізінде құрылады. 1-кестеде көрсетілгендей, әрбір нысан түрі белгілі бір кіріс айнымалыларын қамтиды.

Біз әртүрлі рецензияланған мақалалардың сипаттамалары әртүрлі екенін анықтадық, бұл мұғалімдерге "қара жәшік" үлгілері туралы нақты түсінік алуды қиындатады. XAI әдістері сипаттамалардың үлесін нақтылау және мағыналы интерпретацияны қолдау үшін қажет.

Зерттеулердің 29,7% пайдаланылған мінез-құлық деректері басу және қатысу деректері сияқты айнымалыларды қамтиды. Мысалы, Swamy және басқалары (Суэми және т.б., 2023: 345–356) Scala бағдарламалау курсының сәтсіздікке ұшырау қаупін бейне мен викторинаны көруді қадағалау арқылы болжады. Вахид және басқалары (Уахид және т.б., 2023) Moodle-мен өзара әрекеттесу негізінде STEM курсынан өту немесе сәтсіздік нәтижелерін болжады.

Зерттеулердің 27,0% қолданылатын академиялық үлгерім туралы мәліметтер, мысалы, бағалау нәтижелері және курстың аяқталу мәртебесі нақты жетістік көрсеткіштерін береді. Мысалы, Аднан және басқалары (Аднан және т.б., 2022: 129843–129864) оқушылардың нәтижелерін келесідей жіктеу үшін бағаларды қадағалады: ерекшелену, емтиханнан өту, сәтсіздікке ұшырау немесе кету.

Демографиялық деректер (16,2%) жас және жыныс сияқты айнымалылармен кеңірек контекст береді. Бағдарламалау деректері (13,5%) код пен синтаксистегі қателер сияқты айнымалыларды қамтиды. Мәтінмәндік деректер (13,5%) курстың күрделілігі, ұпайлар және сабақ схемалары сияқты айнымалыларды қамтиды.

Олардың ішінде мінез-құлық сипаттамалары және өнімділік көрсеткіштері ең көп қолданылатын болып ғана емес, сонымен қатар XAI қосымшаларында түсіндіру үшін ең маңызды болып табылады, өйткені олар оқушылардың қатысуы мен жетістіктерін тікелей көрсетеді.

Кесте 2. Кіріс айнымалыларының түрлері [Types of input variables]

Түрі	**Кіріс айнымалылары**	**Пайыз**	**Авторлар**
Мінез-құлықтық	Курс бойынша кликстрим деректері, мәтіндік пікірлер, қатысу, қатысушылық	29,7%	Swamy және т.б., Alrajhi және т.б., Rico-Juan және т.б.
Академиялық үлгерім	ОБК, курсты аяқтау мәртебесі, бағалау балдары	27,0%	Lu және т.б., Giannakas және т.б.
Демографиялық	Жынысы, жасы, ең жоғары білім деңгейі	16,2%	Baranyi және т.б., Mendez және т.б.
Бағдарламалау деректері	Код, пернелерді басу кідірісі, бағдарламалау конструкциялары, синтаксис қателері	13,5%	Pereira және т.б., Waheed және т.б.
Контексттік	Курстың күрделілігі, кредиттер, белсенділік үлгілері	13,5%	Rahul және Katarya, Giannakas және т.б.

Болжау мақсаттары мен алгоритмдері. Болжау мақсаттарын түсіну өте маңызды, өйткені олар XAI әдістері түсіндіруге тырысатын контекстті қамтамасыз етеді. 2-кестеде көрсетілгендей, мақсаттар мен болжау алгоритмдері арасында нақты байланыс бар.

Ең көп таралған мақсат - курстың сәтсіздікке ұшырау қаупі, өйткені зерттеулердің 61,5% студенттердің емтихан тапсыратынын, сәтсіздікке ұшырайтынын немесе оқудан бас тартатынын болжайды. Random Forest (Ригатти, 2017: 31–39), XGBoost (Чен, Гестрин, 2016: 785–794) және LSTM (Грефф және тб., 2016: 2222–2232) сияқты жіктеу алгоритмдері ең көп қолданылады. Мысалы, Лу және оның авторлары (Лу және тб., 2023: 1–9) STEM курсында xgboost көмегімен оқуды болжауда 90% дәлдікке қол жеткізді.

Кесте 3. Студенттердің үлгерімін болжауда қолданылған XAI техникаларына шолу [Review of XAI techniques used in student performance prediction]

Авторлар	Болжау мақсаты	Мақсатты айнымалылар	Үздік алгоритмдер	XAI техникасы	Деңгей	Визуализация әдістері
Swamy және т.б.	Курстан қалу қаупі	Scala бағдарламалауда: өтті/өтпеді	BiLSTM, дәлдігі 90%	SHAP, LIME	Жаһандық	XAI ұқсастық картасы; белгілер маңыздылығының картасы
Baranyi және т.б.	Курстан қалу қаупі	STEM курсын бітірді/шықты	FCNN, дәлдігі 72,4%	SHAP, Permutation Importance	Жаһандық	SHAP жиынтық диаграммасы; маңыздылық кестесі
Alrajhi және т.б.	Студент белсенділігі	Big Data курсындағы пікір кезектілігі	BERT, дәлдігі 92%	Integrated Gradients	Жеке	VisualizationDataRecord (Captum [36])
Mendez және т.б.	Баға немесе ұпай	CS курсындағы баға	GBT (дәлдік көрсетілмеген)	SHAP	Жеке	Дербес дөңгелек диаграммалар (SHAP мәндері)
Adnan және т.б.	Курстан қалу қаупі	STEM: үздік/өтті/ өтпеді/шықты	Random Forest, дәлдігі 91,9%	SHAP	Жаһандық, Жеке	Жаһандық: жиынтық, тәуелділік диаграммасы; Жеке: жергілікті түсіндіру
Pereira және т.б.	Курстан қалу қаупі	Python курсында: өтті/өтпеді	XGBoost, дәлдігі 81,3%	SHAP	Жаһандық, Жеке	Жаһандық: бағаналы, жиынтық; Жеке: шешім, күш диаграммасы
Rahul және Katarya	Курстан қалу қаупі	STEM курсында: өтті/өтпеді	EDL_KDF, дәлдігі 99,6%	SHAP	Жаһандық	Бағаналы диаграмма, жиынтық диаграмма
Rico-Juan және т.б.	Курстан қалу қаупі	Бағдарламалау тапсырмасы: өтті/өтпеді	Random Forest, AUC 0,70	SHAP	Жаһандық	Жиынтық диаграмма, тәуелділік диаграммасы
Lu және т.б.	Курстан қалу қаупі	STEM курсынан шықты/шықпады	XGBoost, дәлдігі 90%, AUC-ROC 0,89	SHAP	Жаһандық, Жеке	Жаһандық: жиынтық, бағаналы, тәуелділік, өзара әрекет, SHAP картасы; Жеке: сарқырама диаграммасы
Giannakas және т.б.	Баға немесе ұпай	Бағдарламалық инженерия: A немесе F	DNN, дәлдігі 82,39%	SHAP	Жаһандық	Бағаналы диаграмма, жиынтық диаграмма
Waheed және т.б.	Курстан қалу қаупі	STEM курсында: өтті/өтпеді	LSTM, дәлдігі 84,57%, AUC 0,82	SHAP	Жаһандық	Жиынтық диаграмма, тәуелділік диаграммасы
Effenberger және Pelánek	Кодты түсіну	Python курсындағы код кластерлері	Дербес кластерлеу алгоритмі	Түсіндірілетін кластерлеу алгоритмі	Жаһандық	Кластерленген белгілер матрицасы, кластер жиынтық диаграммасы
Afzaal және т.б.	Баға немесе ұпай	Бағдарламалау тапсырмаларының нәтижелері	ANN, дәлдігі 90%	Counterfactual Explanations	Жеке	Кері байланыс бақылау тақтасы (DICE)

ISSN (print) 3135-1840 eISSN (online) 3135-1859 Smart Technologies Journal, 2(2) · 2026

3-кестеде зерттеулердің 23,1% нәтижелерін болжау үшін gradientboosting Trees (GBT) (Фридман, 2001: 1189–1232), жасанды нейрондық желілер (ANN) (Егнанараяна, 2009) және терең нейрондық желілер (DNN) (Ше және т.б., 2017: 2295–2329) сияқты күшейтетін алгоритмдер немесе нейрондық желілер қолданылды. Мысалы, Мендес және басқалар. (Мендес және т.б., 2021) информатика курсының тұрақты бағаларын болжау үшін регрессия GBT қолданды. "Кодты түсіну" санатында Эфенбергер мен Пеланек (Эффенбергер және Пеланек, 2021: 101–112) Python курсында код ерекшеліктеріне негізделген студенттердің бағдарламалық шешімдерін топтастыру үшін кластерлеуді қолданды. Студенттерді Тарту үшін Алраджи және т.б. (Әлражхи және т.б., 2023) трансформаторлардан Екі Бағытты Кодтаушы Көріністері бар қолданбалы мәтінді өндіру (Девлин, 2018) (BERT негізіндегі нейрондық желі) жеделдігін жіктеу үшін студент Big Data курсында түсініктеме беріп, 92% жоғары дәлдікке қол жеткізді.

Rq2: XAI әдістері, интерпретация деңгейлері, бейнелеу әдістері және практикалық іске асыру. Болжамды модельдердің интерпретациясын жақсарту үшін әртүрлі XAI әдістері қолданылды. Ең көп қолданылатын әдіс ретінде Шеплидің қосымша түсіндірмелері (SHAP) (Лундберг, Ли, 2017) 11 зерттеуде келтірілген. Мысалы, Перейра (Перейра және т.б., 2021: 117097–117119) XGBOOST моделімен SHAP-ті оқушылардың негізгі мінез-құлқын анықтау үшін қолданды, мұғалімдерге табысқа қандай факторлар ықпал ететінін түсінуге көмектесті және Python курсында тәуекел тобындағы студенттерге мақсатты қолдау көрсетті. SHAP-тің танымалдылығын оның ойын теориясындағы берік теориялық негізімен (Шепли, 1953), жергілікті және жаһандық түсініктемелер беру қабілетімен, машиналық оқыту модельдерінің көпшілігімен жұмыс істеу икемділігімен және ашық бастапқы кітапханалармен қол жетімділігімен түсіндіруге болады.

LIME. Тек бір зерттеуде жергілікті түсіндірілетін модельдік диагностикалық түсініктемелер (LIME) қолданылды (Рибейро и др., 2016: 1135–1144), әр жағдайға жеңілдетілген модельдер жасау арқылы жеке болжамдарды түсіндіретін кеңінен қолданылатын әдіс. LIME нақты болжамдар үшін күрделі модельдің мінез-құлқын жуықтау үшін сызықтық регрессиялар немесе шешім ағаштары сияқты жергілікті суррогат модельдерін жасайды. Атап айтқанда, Swamy (Суэми және т.б., 2023: 345–356) викториналар жіберу және бейнелерге қатысу сияқты сәттілікке әсер ететін маңызды факторларды анықтау үшін LIME қолданды.

Ауыстырудың маңыздылығы. Зерттеулердің бірінде Permutation Importance (Альтман және т.б., 2010: 1340–1347) әдісі қолданылды, ол әр белгінің мәндерін араластыру және модельдің дәлдігіне әсерін өлшеу арқылы белгілердің маңыздылығын бағалауға мүмкіндік береді. Дәлдіктің айтарлықтай төмендеуі маңызды белгіні көрсетеді. Бараньи және басқалар (Бараньи және т.б., 2020: 13–19) оны STEM курсының маңызды көрсеткіштері ретінде университетке түсу кезінде балл сияқты факторларды растай отырып, бітіруді болжаушыларды тексеру үшін пайдаланды.

Біріктірілген градиенттер. Альраджи және басқалар (Альраджхи және т.б., 2023) оқушылардың шұғыл түсініктемелерін алу үшін Bert моделін түсіндіру үшін объектілерге апаратын жолдар бойындағы градиенттерді есептейтін (Сундарараджан және т.б., 2017: 3319–3328) интеграцияланған градиенттерді қолданды. Бұл әдіс мұғалімдерге жедел жауап беруді қажет ететін аймақтарды анықтауға және модельдің ашықтығын арттыруға көмектесетін кілт сөздерді бөліп көрсетті.

Контрафактілік Түсініктемелер. Афзаалда және басқалар (Афзаал және т.б., 2021: 37–42), контрафактілік түсініктемелер (Вахтер және т.б., 2017: 841) мүмкіндік мәндерін реттеу арқылы "what-if" сценарийлерін жасау үшін пайдаланылды. Бұл тәсіл енгізу мүмкіндіктеріндегі ең аз өзгерістердің модельдік болжамдарға қалай әсер ететінін зерттейді. Бұл зерттеу бағдарламалау курсында студенттердің мүмкіндіктерін кеңейту үшін жекелендірілген ұсыныстарды пайдалануға бағытталған, бұл олардың үлгерімін жақсарту үшін мақсатты әрекеттерді орындауға мүмкіндік береді.

Пайдаланушы түсіндіретін кластерлеу алгоритмі. Эфенбергер және Пеланек (Эффенбергер, Пеланек, 2021: 101–112) Python курсында оқушылар ұсынған кодты

жіктеу үшін пайдаланушының интерпретацияланған кластерлеу алгоритмін цикл және шартты операторлар сияқты стандартты бағдарламалау үлгілерінде пайдаланды. Бұл тәсіл мұғалімдерге жалпы кодтау әдістерін жылдам анықтауға мүмкіндік береді, бұл мақсатты кері байланысты жеңілдетеді және бағдарламалау тапсырмаларындағы өнімділік деректерінің интерпретациясын арттырады. Әр түрлі XAI әдістері қолданылғанымен, қарастырылған зерттеулерде жинақталған жергілікті эффекттерде (ALE) (Апли, Жу, 2020: 1059–1086) байланыстыру (Рибейро және т.б., 2018) және терең оқытудың маңызды функциялары (DeepLIFT) (Шрикумар және т.б., 2017: 3145–3153) сияқты әдістер айтарлықтай жетіспеді. Бұл алшақтық болашақ зерттеулерге осы әдістерді оқушылардың үлгерімін болжау модельдеріне енгізу мүмкіндігін ашады. Біз XAI-дің информатика және STEM білім берудегі практикалық қосымшаларына арналған әдебиеттерге бірде-бір шолуды таппадық. Бұл оқылықты жою үшін біз 1-кестеде білім беру нәтижелеріне ықпал ететін тиісті қосымшаларды қорытындыладық.

Кесте 4. CS және STEM білімінде XAI қолдану [Application of XAI in CS and STEM education]

Санат	**Пайыз**
Жаһандық белгілердің маңыздылығын талдау	33,3%
Жеке болжамды түсіндіру	20%
Араласуларды және шешім қабылдауды қолдау	26,7%
Белгілердің өзара әрекеттесуін зерттеу	13,3%
Курсты жақсартуға бағытталған іс-шаралар	3,3%
Ұсынымдар жасау	3,3%

Жаһандық белгілердің маңыздылығын талдау. Қарастырылған зерттеулердің 33,3% құрайтын негізгі қолдану жаһандық белгілердің маңыздылығын талдау болып табылады, оның мақсаты оқушылардың жалпы үлгерімі үшін ең маңызды белгілерді анықтау болып табылады. Мысалы, Перейра және бірлескен авторлар (Перейра және т.б., 2021: 117097–117119) студенттердің үлгеріміне әсер ететін маңызды факторларды талдау үшін SHAP жиынтық кестелерін қолданды және бірінші емтихан бағасы мен жүйеге кіруге кететін уақыт CS1 курсының оқу үлгерімінің негізгі факторлары екенін анықтады.

Жеке болжамды түсіндіру. Зерттеулердің 20% - ы жеке болжамды түсіндіру арқылы жеке оқушылардың мінез-құлық үлгілеріне әсер ететін, мақсатты ақпараттарды бөліп көрсетеді. Мысалы, Лу (Лу және т.б., 2023: 1–9) ШАП сарқырамасының кестесін (SHAP Waterfall Plot) ресурстар мен тапсырмаларды басу сияқты функциялардың әсерін көрсету үшін және бұл факторлар оқушының STEM курсында оқуын жалғастыру немесе оны тастап кету ықтималдығына қалай әсер ететінін түсіндіру үшін пайдаланды.

Қолдау Шаралары Және Шешім қабылдау. Зерттеулердің 26,7% - ында XAI қосымшалары оқытушыларға оқушылардың үлгерімін қолдау үшін уақытылы араласуға көмектеседі. Мысалы, Алраджи және басқалар (Альраджхи және т.б., 2023) Captum визуализациясын (Кохликян және т.б., 2020) мұғалімдерге шұғыл назар аударуды қажет ететін түсініктемелерді анықтауға көмектесу үшін, қажет болған жағдайда мақсатты қолдау көрсетуге мүмкіндік беру үшін қолданды.

Ерекшеліктердің Өзара Әрекеттесуін Зерттеу. Зерттеулердің 13,3% -ын құрайтын функционалдық өзара әрекеттесуді зерттеу әртүрлі факторлардың оқушылардың үлгеріміне қалай әсер ететінін зерттейді. Мысалы, Вахид және басқалар (Уахид және т.б., 2023) викторинаға қатысу мен курстық тапсырмалар арасындағы байланысты зерттеу үшін SHAP тәуелділік сюжеттерін пайдаланды. Олар екі іс-шараға да дәйекті қатысу сәтсіздік қаупі төменгі деңгеймен байланысты екенін анықтады, бұл жиынтық оқушылардың үлгеріміне жиынтық оң әсерін көрсететінін анықтады.

Курсты практикалық жетілдіру. Қарастырылған зерттеулердің біреуінде ғана ХАІ курстың жақсаруын іздеу үшін пайдаланылды. Свами және басқалары (Суэми және т.б., 2023: 345–356) сапалы сауалнамалар жүргізіп, оқытушылардың 85% - дан астамы викториналық тапсырмаларды түзету сияқты бағдарламалау курсының элементтерін нақтылау үшін ХАІ идеяларын бағалағанын анықтады.

Ұсыныстар Жасау. Сондай-ақ, қарастырылған зерттеулердің бірінде жекелендірілген кері байланысқа бағытталған ұсыныстар жасалды. Афзаал және басқалары (Афзаал және т.б., 2021: 37–42) арнайы кеңестер ұсынатын бақылау тақтасын жасады. Мысалы, егер викториналық ұпайлар жақсартуды ұсынса, онда тақтасы студенттерге тиісті оқу материалдары бойынша ұсыныстар жасайды, бұл студенттерге бағдарламалау дағдыларын жетілдірудің практикалық стратегияларына назар аударуға көмектеседі.

Осылайша, жекелендірілген ұсыныстар жасау үшін ХАІ пайдалануда айтарлықтай алшақтық бар. Бұл олқылықты жою оқытушыларға жеке студенттерді жақсырақ қолдау үшін жекелендірілген, деректерге негізделген ұсыныстар ұсынуға мүмкіндік береді. Сонымен қатар, ХАІ курсын жетілдіру бойынша ұсыныстар шектеулі болып қала береді, бұл оқытуды жақсарту және оқу нәтижелерін оңтайландыру үшін қосымша мүмкіндіктер ашады.

Түсіндіру деңгейлері. Оқушылардың оқу үлгерімін болжау үшін ХАІ әдістерін зерттеген кезде, түсіндіру әдістері жаһандық және жеке (жергілікті) деңгейлерге жіктеледі, олардың әрқайсысы ерекше түсініктер ұсынады. Қарастырылған жеті зерттеу жаһандық деңгейдегі түсіндірмелерге, үшеуі жеке деңгейдегі түсіндірмелерге бағытталған, ал үшеуі екеуін де қамтиды.

Әлемдік деңгей. Әлемдік түсіндірме барлық болжамдар бойынша мүмкіндіктердің маңыздылығы туралы түсінік береді, бұл деректер жиынтығы бойынша үлгілерді бақылауды жеңілдетеді. Қарастырылған зерттеулерде SHAP әдетте мүмкіндіктердің маңыздылығы мен тәуелділіктерін әлемдік деңгейде анықтау үшін қолданылады. Ол жан-жақты түсініктер бергенімен, SHAP жоғары есептеу ресурстарын, әсіресе жоғары өлшемді деректер үшін қажет, бұл өңдеу шығындарын арттыруы мүмкін.

Жеке деңгей. Жеке түсіндіру мүмкіндігі жеке болжамдарға ерекшеліктердің әсеріне бағытталған, жеке оқушыларға арналған жекелендірілген талдауларды қолдайды. Интеграцияланған градиенттер (Альраджхи и др., 2023) және контрафактілік түсіндірмелер (Афзаал және т.б., 2021: 37–42) сияқты әдістер ерекшеліктерге негізделген үлестер мен нәтижеге әсер ететін минималды өзгерістерді қамтамасыз етеді, бұл жеке араласуларға мүмкіндік береді.

Екінші жағынан, LIME шолудан өткен зерттеулердің бірінде пайда болды (Суэми және т.б., 2023: 345–356). LIME болжамды жақындату үшін жергілікті суррогат моделін жасау арқылы SHAP-қа қарағанда жеке түсіндірмелердің интуитивті визуализациясын ұсынады. Бұл пайдаланушыларға әрбір мүмкіндіктің белгілі бір жағдай үшін болжамды нәтижеге қалай үлес қосатынын жақсы түсінуге көмектеседі.

Қарастырылған зерттеулердің көпшілігі жаһандық түсіндірулерге баса назар аударады, дегенмен жеке деңгейдегі талдау жеке тұлғалар, білім беру араласулары үшін шешуші болып қала береді. Сонымен қатар, Свами және басқалар (Суэми және т.б., 2023: 345–356) атап өткендей SHAP және LIME сияқты әртүрлі ХАІ әдістері бір деректер жиынтығында әртүрлі түсініктемелер бере алады. Бұл мәселелер зерттеушілер неліктен қолдау үшін бір ХАІ әдісін қолданғанын түсіндіре алады түсініктемелердің дәйектілігі. Бұл мәселелер зерттеушілердің түсіндірмелердің дәйектілігін сақтау үшін ХАІ әдісін не үшін қолданғанын түсіндіруі мүмкін. Өнімділікті болжау модельдерін жасау кезінде болжау дәлдігі мен интерпретация арасындағы тепе-теңдікті қамтамасыз ету маңызды.

Кесте 5. Машиналық оқыту және терең оқыту [Machine learning and deep learning]

Түрі	Пайыз	Алгоритм	Модель дәлдігі	Үздік алгоритм
Дәстүрлі машиналық оқыту	53,8%	GBT, ANN, Random Forest, XGBoost, Custom clustering	88,3% ± 4,8	Random Forest, дәлдігі 91,9%
Терең оқыту	46,2%	BiLSTM, FCNN, BERT, EDL_KDF, DNN, LSTM	86,8% ± 9,3	EDL_KDF, дәлдігі 99,6%

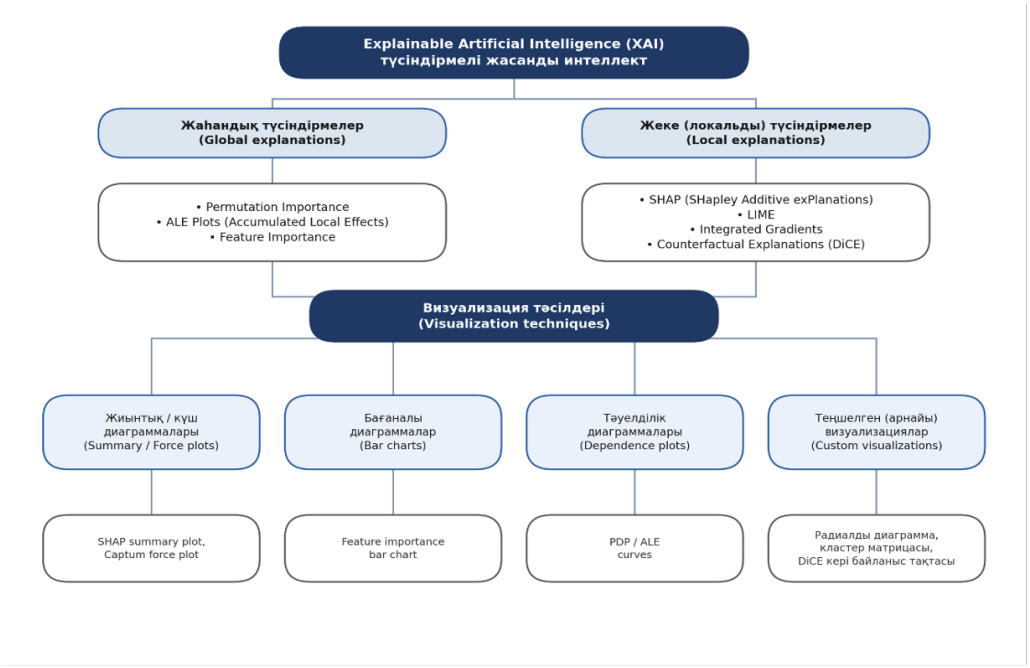
Дәстүрлі машиналық оқыту. 5-кестеде көрсетілгендей, GBT, Random Forest, XGBoost, ANN және арнайы кластерлеу алгоритмдері қарастырылып отырған тәсілдердің 53,8%-ын құрайды. Бұл дәстүрлі машиналық оқыту модельдері орташа дәлдікті 88,3 ± 4,8% құрайды, ал кездейсоқ орман 91,9% ең жақсы нәтижелерге қол жеткізеді (Аднан және т.б., 2022: 129843–129864). Бұл модельдер түсіндіруді жақсарту үшін SHAP және контрафактілік түсіндірмелер сияқты XAI әдістерін пайдаланады.

Терең оқыту. Екі бағытты LSTM (BiLSTM), толықтай байланысты нейрондық желілер (FCNN), BERT, айқын терең оқыту білімін бөлу фреймворкі (EDL_KDF), DNN және LSTM сияқты озық архитектуралар әдістердің 46,2%-ын құрайды және орташа дәлдік 86,8% ± 9,3%-ға жетеді. Бұл өзгергіштік терең оқытудың одан да жоғары өнімділікті қамтамасыз ету мүмкіндігі бар екенін көрсетеді. Терең оқыту дәстүрлі түрде онша түсіндірілмейтін болып саналса да (Мин және т.б., 2022: 3503–3568), бұл шолу терең оқыту модельдерін түсіндіру үшін SHAP, LIME және интеграцияланған градиенттер сияқты әртүрлі XAI әдістерін пайдалануға болатынын анықтады. Терең оқыту моделі As DAbDBi-LSTM және ResNet-101 желілерін біріктірген EDL_KDF терең оқыту құрылымын (Рахул, Катарйа, 2024: 1–8) пайдаланып, ең жоғары 99,6% дәлдікке қол жеткізді. Бұл құрылым дәстүрлі машиналық оқыту модельдерінен асып түсті. Сондықтан, егер болжау дәлдігі негізгі мақсат болса және терең оқыту модельдерінің түсіндірілу мәселелерін шешуге болатын болса, терең оқыту алгоритмдері бірінші таңдау болуы керек.

Оқушылардың үлгерімін болжауға арналған XAI визуализация әдістерін екі негізгі санатқа бөлуге болады.

XAI арқылы жасалған графиктер. Бұл тәсіл кітапханалардан алынған стандартты графиктерді пайдаланып, модель болжамдарына ерекшеліктердің әсерін визуализациялады, бұл оқытушылар мен зерттеушілерге күрделі модельдерді түсінуге көмектеседі. Альраджхи және басқалары (Альраджхи және т.б., 2023) Captum интеграцияланған градиенттерін пайдаланып, түсініктемелерде жеделдік жіктемелеріне әсер ететін кілт сөздерді белгіледі. Сонымен қатар, Baranyi және басқалары (Бараньи және т.б., 2020: 13–19) нақты факторлардың бітіру ықтималдығына қалай әсер ететінін көрсету үшін SHAP қорытынды графиктерін пайдаланды. Қарастырылған мақалалар sci-kit-learn (Рико-Хуан және т.б., 2023: 955–969), TensorFlow (Суэми және т.б., 2023: 345–356) және PyTorch (Афзаал және т.б., 2021: 37–42) бағдарламаларындағы XAI кітапханаларын пайдаланып жасалған визуализациялардың түсіндірілетін шешімдерді жылдамырақ әзірлеуге мүмкіндік беретінін көрсетті.

Теңшелген визуализациялар. Бұл тәсіл XAI нәтижелерін қолжетімділікті жақсарту үшін теңшейді, бұл оқытушылар мен оқушылардың түсіндіруін жеңілдетеді. Мендес және т.б. (Мендес және т.б., 2021) ерекшелік үлестерін көрнекі түрде нақтылау үшін SHAP мәндеріне негізделген дөңгелек диаграммалар жасады. Дегенмен, бұл бағытты зерттеген зерттеулер аз, бұл оқытушы мен оқушыға ыңғайлы визуализация құралдарын әзірлеудегі алшақтықты көрсетеді. Мұны шешу білім беру мекемелерінде XAI әдістерінің қолжетімділігін және практикалық құндылығын арттыруы мүмкін.



Сурет 2. Explainable Artificial Intelligence (XAI) әдістері мен визуализация тәсілдерінің жіктелуі [Classification of Explainable Artificial Intelligence (XAI) methods and visualization techniques]

Зерттеудің шектеулері (Limitations). Осы жүйелі шолудың нәтижелерін интерпретациялау кезінде бірнеше әдістемелік шектеулерді ескеру қажет. Біріншіден, әдебиеттерді іздеу тек таңдалған негізгі ғылыми дерекқорлармен (ACM Digital Library, IEEE Xplore және Web of Science) шектелді. Сонымен қатар, талдауға тек ағылшын тіліндегі басылымдар ғана енгізілді, бұл басқа тілдердегі құнды зерттеулердің ескерілмеуіне байланысты «тілдік ауытқуға» (language bias) алып келуі ықтимал. Екіншіден, қосу және алып тастаудың қатаң критерийлеріне сәйкес, шолуға іріктелген зерттеулердің саны шектеулі, бұл алынған қорытындыларды жалпылауға белгілі бір шектеулер қояды. Түсіндірмелі жасанды интеллект (XAI) саласының өте қарқынды дамып жатқанын ескерсек, іздеу процесі аяқталғаннан кейін жарық көрген ең жаңа немесе инновациялық еңбектер осы шолуға енбей қалуы мүмкін. Сонымен қатар, зерттеудің тек STEM және компьютерлік ғылымдар саласына бағытталғандығы алынған қорытындыларды басқа гуманитарлық немесе әлеуметтік пәндерге тікелей қолдану мүмкіндігін шектейді.

Қорытынды. Бұл жүйелі әдебиеттік шолуда STEM пәндері мен компьютерлік ғылымдар аясында үлгерімді болжау модельдерін интерпретациялау үшін түсіндірмелі жасанды интеллект (XAI) құралдарын қолданудың қазіргі жағдайы зерттелді.

Зерттеудің негізгі нәтижелері келесідей тұжырымдалды:

Деректер мен белгілерді талдау (RQ1): Жұмыстардың басым бөлігі тікелей оқытуды басқару жүйелерінен (LMS) алынған авторлық деректер жиынтығына негізделген. Студенттердің мінез-құлық үлгілері мен академиялық көрсеткіштері ең жиі қолданылатын белгілер болып табылады, ал болжамның басты мақсаты — үлгермеушілік қаупін анықтау. Қолданылатын белгілердің әртүрлілігі педагогтар үшін «қара жәшік» модельдерінің логикасын түсінуді қиындатады, бұл XAI әдістерін енгізу қажеттілігін арттырады.

XAI-ды қолдану салалары (RQ2): Әдістер көбінесе белгілердің жаһандық маңыздылығын талдау, жеке болжамдарды түсіндіру және педагогикалық шешімдерді қабылдауды қолдау үшін пайдаланылады. SHAP әдісі ең көп таралған құрал ретінде анықталды, бірақ ол негізінен жаһандық деңгейде қолданылып, жеке жағдайларды талдауда шектеулі болып отыр. Сонымен қатар, ALE, Anchors және DeepLIFT сияқты әдістердің әдебиетте кездеспеуі болашақ зерттеулер үшін жаңа мүмкіндіктерді көрсетеді.

Анықталған олқылықтар: XAI-ды оқу курстарын жетілдіру, бейімделген

визуализацияларды жасау және персоналдандырылған ұсыныстар беру үшін қолдануда айтарлықтай тапшылық байқалады. Болжау дәлдігі басымдыққа ие болған жағдайда, терең оқыту (deep learning) модельдері дәстүрлі алгоритмдерге қарағанда жоғары тиімділік көрсетті. Аталған ғылыми олқылықтарды жою білім берудегі XAI-дың практикалық маңызын арттыруға мүмкіндік береді. Бұл оқытушыларға STEM және компьютерлік ғылымдар саласындағы студенттерге неғұрлым тиімді әрі дербес қолдау көрсетуге жағдай жасайды.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Adnan, M., Uddin, M. I., Khan, E., Alharithi, F. S., Amin, S., & Alzahrani, A. A. (2022). Earliest possible global and local interpretation of students' performance in virtual learning environment by leveraging explainable AI. IEEE Access, 10, 129843–129864. https://doi.org/10.1109/ACCESS.2022.3227072

Afzaal, M., et al. (2021). Generation of automatic data-driven feedback to students using explainable machine learning. In International Conference on Artificial Intelligence in Education (pp. 37–42). Springer.

Alrajhi, L., Pereira, F. D., Cristea, A. I., & Alamri, A. (2023). Serendipitous gains of explaining a classifier: Artificial versus human performance and annotator support in an urgent instructor-intervention model for MOOCs. In Proceedings of the 6th Workshop on Human Factors in Hypertext. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3603607.3613480

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340–1347.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), 1059–1086.

Arrieta, A. B., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3–17.

Baranyi, M., Nagy, M., & Molontay, R. (2020). Interpretable deep learning for university dropout prediction. In Proceedings of the 21st Annual Conference on Information Technology Education (pp. 13–19). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3368308.3415382

Chaushi, B. A., Selimi, B., Chaushi, A., & Apostolova, M. (2023). Explainable artificial intelligence in education: A comprehensive review. In World Conference on Explainable Artificial Intelligence (pp. 48–71). Springer.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2023). A systematic literature review on performance prediction in learning programming using educational data mining. In 2023 IEEE Frontiers in Education Conference (FIE).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024a). How various educational features influence programming performance in primary school education. In 2024 IEEE Global Engineering Education Conference (EDUCON) (pp. 1–8). IEEE.

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024b). Enhance learning performance predictions with explainable machine learning. In 2024 IEEE Frontiers in Education Conference (FIE).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024c). Analyzing the interpretability of machine learning prediction on student performance using SHapley Additive exPlanations. In 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) (pp. 1–8). IEEE.

Devlin, J. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Effenberger, T., & Pelánek, R. (2021). Interpretable clustering of students' solutions in introductory programming. In International Conference on Artificial Intelligence in Education (pp. 101–112). Springer.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Giannakas, F., Troussas, C., Voyiatzis, I., & Sgouropoulou, C. (2021). A deep learning classification framework for early prediction of team-based academic performance. Applied Soft Computing, 106, 107355. https://doi.org/10.1016/j.asoc.2021.107355

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 80–89). IEEE.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222–2232.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42.

Kelleher, C., & Pausch, R. (2005). Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. ACM Computing Surveys, 37(2), 83–137.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. UK.

Kokhlikyan, N., et al. (2020). Captum: A unified and generic model interpretability library for PyTorch. arXiv preprint arXiv:2009.07896.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. Scientific Data, 4(1), 1–8.

Lu, J., Mou, J., & Li, P. (2023). Interpretive analyses of learner dropout prediction in online STEM courses. In 2023 5th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 1–9). https://doi.org/10.1109/CSTE59648.2023.00020

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30).

Marcinkevičs, R., & Vogt, J. E. (2020). Interpretability and explainability: A machine learning zoo mini-tour. arXiv preprint arXiv:2012.01805.

Mathrani, A., Susnjak, T., Ramaswami, G., & Barczak, A. (2021). Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. Computers and Education Open, 2, 100060.

Mendez, G., Galárraga, L., & Chiluiza, K. (2021). Showing academic performance predictions during term planning: Effects on students' decisions, behaviors, and preferences. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445718

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. Artificial Intelligence Review, 55(5), 3503–3568. https://doi.org/10.1007/s10462-021-10088-y

Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 607–617).

Pereira, F. D., et al. (2021). Explaining individual and collective programming students' behavior by interpreting a black-box predictive model. IEEE Access, 9, 117097–117119. https://doi.org/10.1109/ACCESS.2021.3105956

Rahul, & Katarya, R. (2024). Shapley explainable deep learning based knowledge distillation framework for student's performance prediction. In 2024 Second International Conference on Data Science and Information System (ICDSIS) (pp. 1–8). https://doi.org/10.1109/ICDSIS61070.2024.10594386
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).
Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI Conference on Artificial Intelligence.
Rico-Juan, J. R., Sánchez-Cartagena, V. M., Valero-Mas, J. J., & Gallego, A. J. (2023). Identifying student profiles within online judge systems using explainable artificial intelligence. IEEE Transactions on Learning Technologies, 16(6), 955–969. https://doi.org/10.1109/TLT.2023.3239110
Rigatti, S. J. (2017). Random forest. Journal of Insurance Medicine, 47(1), 31–39.
Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368–384.
Shapley, L. S. (1953). A value for n-person games. Princeton University Press.
Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning (pp. 3145–3153). PMLR.
Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning (pp. 3319–3328). PMLR.
Swamy, V., Du, S., Marras, M., & Kaser, T. (2023). Trusting the explainers: Teacher validation of explainable artificial intelligence for course design. In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 345–356). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3576050.3576147
Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295–2329
Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31, 841.
Waheed, H., Hassan, S.-U., Nawaz, R., Aljohani, N. R., Chen, G., & Gasevic, D. (2023). Early prediction of learners at risk in self-paced education: A neural network approach. Expert Systems with Applications, 213(A), 118868. https://doi.org/10.1016/j.eswa.2022.118868
Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd.

Редактор: Мырзабекова А.М. Верстка: Сексенова Ж.М. Подписано в печать: 30.06.2026 г. Издание: ТОО Международный университет Астана 010000, Казахстан, г. Астана, пр. Кабанбай батыра, 8, тел.: +7 (7172) 47-62-10 (214), e-mail: stj@aiu.edu.kz